



US009243198B2

(12) **United States Patent**
Reppas et al.

(10) **Patent No.:** **US 9,243,198 B2**
(45) **Date of Patent:** **Jan. 26, 2016**

(54) **METHODS AND COMPOSITIONS FOR THE RECOMBINANT BIOSYNTHESIS OF N-ALKANES**

(71) Applicant: **Joule Unlimited Technologies, Inc.**,
Bedford, MA (US)

(72) Inventors: **Nikos Basil Reppas**, Brookline, MA
(US); **Christian Perry Ridley**, Acton,
MA (US)

(73) Assignee: **Joule Unlimited Technologies, Inc.**,
Bedford, MA (US)

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 0 days.

(21) Appl. No.: **14/593,255**

(22) Filed: **Jan. 9, 2015**

(65) **Prior Publication Data**

US 2015/0284649 A1 Oct. 8, 2015

Related U.S. Application Data

(60) Division of application No. 13/442,840, filed on Apr.
9, 2012, now Pat. No. 8,932,872, which is a
continuation of application No. 13/243,136, filed on
Sep. 23, 2011, now Pat. No. 8,183,027, which is a
continuation of application No. 13/098,700, filed on
May 2, 2011, now Pat. No. 8,043,840, which is a
continuation of application No. 12/833,821, filed on
Jul. 9, 2010, now Pat. No. 7,955,820, which is a
continuation-in-part of application No. 12/759,657,
filed on Apr. 13, 2010, now Pat. No. 7,794,969.

(60) Provisional application No. 61/224,463, filed on Jul. 9,
2009, provisional application No. 61/228,937, filed on
Jul. 27, 2009.

(51) **Int. Cl.**
G01N 33/00 (2006.01)
C10L 1/04 (2006.01)
C10L 1/08 (2006.01)

(52) **U.S. Cl.**
CPC ... **C10L 1/04** (2013.01); **C10L 1/08** (2013.01);
C10L 2200/0469 (2013.01); **C10L 2270/026**
(2013.01)

(58) **Field of Classification Search**
CPC C10L 1/04
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,223,876 B2 5/2007 Gobel et al.
7,794,969 B1 9/2010 Reppas et al.
7,919,303 B2 4/2011 Reppas et al.
7,955,820 B1 6/2011 Reppas et al.
8,043,840 B2 10/2011 Reppas et al.
8,101,397 B2 1/2012 Reppas et al.
8,183,027 B2 5/2012 Reppas et al.
8,481,285 B2 7/2013 Reppas et al.

2004/0048343 A1 3/2004 Hermann et al.
2006/0225145 A1 10/2006 Ozawa
2009/0004715 A1 1/2009 Tribmur et al.
2009/0011480 A1 1/2009 Tribmur et al.
2009/0061493 A1 3/2009 Tribmur et al.
2009/0155873 A1 6/2009 Kashiyaama
2009/0203070 A1 8/2009 Devroe et al.
2010/0136595 A1 6/2010 Tittiger et al.
2010/0151539 A1 6/2010 Franklin et al.
2010/0154293 A1 6/2010 Hom et al.
2010/0170826 A1 7/2010 Friedman et al.
2010/0221798 A1 9/2010 Schirmer et al.
2010/0249470 A1 9/2010 Schirmer et al.
2012/0135486 A1 5/2012 Reppas et al.

FOREIGN PATENT DOCUMENTS

CN 102015995 A 4/2011
CN 102027108 A 4/2011
CN 102027109 A 4/2011
EP 711351 B1 3/2006
EP 2 305 778 A2 4/2011
JP H07-300593 A 11/1995
WO WO 92/14816 A1 9/1992
WO WO 2007/136762 11/2007
WO WO 2008/003078 1/2008
WO WO-2008/119082 10/2008
WO WO 2008/147781 A2 12/2008
WO WO 2009/006430 1/2009
WO WO 2009/036095 A1 3/2009
WO WO 2009/062190 5/2009
WO WO 2009/111513 9/2009
WO WO 2009/140695 11/2009
WO WO 2009/140696 11/2009
WO WO 2010/068288 6/2010

OTHER PUBLICATIONS

Australian Government IP Australia, Patent Examination Report No.
1, Australian Patent Application No. 2011-204 785. Oct. 8, 2012,
three pages
Australian Patent Office. Examiner's First Report. Patent Application
No. 2010-246473, Jan. 24, 2011, two pages.
Beer, I. L. et al., "Engineering Algae for Biohydrogen and Biofuel
Production," Current Opinion in Biotechnology, 2009, vol. 20, No. 3,
pp. 264-271.
Canadian Intellectual Property Office, Office Action, Canadian
Patent Application No. 2,766,204, Jul. 31, 2013, two pages.

(Continued)

Primary Examiner — Albert Navarro

Assistant Examiner — Mark Navarro

(74) *Attorney, Agent, or Firm* — Fenwick & West LLP;
Chang B. Hong, Esq.

(57) **ABSTRACT**

The present disclosure identifies methods and compositions
for modifying photoautotrophic organisms as hosts, such that
the organisms efficiently convert carbon dioxide and light
into n-alkanes, and in particular the use of such organisms for
the commercial production of n-alkanes and related mol-
ecules.

(56)

References Cited

OTHER PUBLICATIONS

- Copeland, A. et al. UniProt Direct Submission ID O31 C01, Jan. 24, 2006, p. 1 [Online] [Retrieved Dec. 7, 2010] Retrieved from the Internet URL:<http://www.uniprot.org/uniprot/Q31C01.txt?version=1>>.
- Copeland, A. et al. UniProt Direct Submission ID O31C02, Jan. 24, 2006, p. 1 [Online] [Retrieved Dec. 7, 2010] Retrieved from the Internet URL:<http://www.uniprot.org/uniprot/Q31C02.txt?version=1>.
- Eurasian Patent Office, Office Action, Eurasian Patent Application No. 201270151, Apr. 19, 2013, four pages.
- European Patent Office, European Search Report and Opinion, European Patent Application No. 13152069.4, Jul. 23, 2013, ten pages.
- European Patent Office, Examination Report, European Patent Application No. EP 10776070.4, Dec. 8, 2011, four pages.
- European Patent Office, Supplementary European Search Report, European Patent Application No. 10776070.4, Aug. 25, 2011, Seven pages.
- European Patent Office, Search Report and Opinion, European Patent Application No. 13197830.6, Sep. 5, 2014, seven pages.
- Fehler, S.W.G. et al., "Biosynthesis of Hydrocarbons in *Anabaena Variabilis*. Incorporation of [methyl-¹⁴C]-and [methyl-²H₃] Methionine into 7- and 8-Methylheptadecanes," *Biochemistry*, 1970, vol. 8, No. 2, pp. 418-422.
- Heipieper, H., "Adaptation of *Escherichia coli* to Ethanol on the Level of Membrane Fatty Acid Composition," *Applied Environmental Microbiology*, Jun. 2005, p. 3388, vol. 71, No. 6.
- Hellingwerf, K.J. et al., "Alternative Routes to Biofuels: Light-driven Biofuel Formation from CO₂ and Water Based on the 'Photanol' Approach," *Journal of Biotechnology*, 2009, pp. 87-90, vol. 142.
- Huber, G.W. et al., "Processing biomass in conventional oil refineries: Production of high quality diesel by hydrotreating vegetable oils in heavy vacuum oil mixtures," *Applied Catalysis A: General*, 2007, pp. 120-129, vol. 329.
- Jetter, R., et al., "Plant surface lipid biosynthetic pathways and their utility for metabolic engineering of waxes and hydrocarbon biofuels." *The Plant Journal*, 2008, pp. 670-683, vol. 54.
- Ladygina, N., et al., "A review on microbial synthesis of hydrocarbons," *Process Biochemistry*, 2006, pp. 1001-1014, vol. 41, No. 5.
- Li, N. et al., "Conversion of Fatty Aldehydes to Alka(e)nes and Formate by a Cyanobacterial Aldehyde Decarbonylase: Cryptic Redox by an Unusual Di metal Oxygenase," *Journal of the American Chemical Society*, Apr. 2011, pp. 6158-6161, vol. 133, No. 16.
- Li, N. et al., "Conversion of Fatty Aldehydes to Alka(e)nes and Formate by a Cyanobacterial Aldehyde C1 Decarbonylase: Cryptic Redox by an Unusual Di metal Oxygenase," *Journal of the American Chemical Society*, Feb. 12, 2011, pp. A-D (doi:10.1021/ia2013517).
- Luque, I. et al. *Synechococcus* sp. nir Gene for Nitrile Reductase. GenBank Direct Submission Accession X67680.1, May 27, 1993, pp. 1-2 [Online] [Retrieved Dec. 7, 2010] Retrieved from the Internet URL:<http://www.ncbi.nlm.nih.gov/nucore/288053>>.
- Maeda, S-I, et al., "cis-Acting Sequences Required for NicB-Dependent, Nitrite-Responsive Positive Regulation of the Nitrate Assimilation Operon in the Cyanobacterium *Synechococcus* sp. Strain PCC 7942," *Journal of Bacteriology*, 1998, pp. 4080-4088, vol. 180, No. 16.
- Mexican Intellectual Property Office, Official Action, Mexican Patent Application No. 12/000448, Aug. 26, 2013, three pages.
- Mpuru, S. et al., "Mechanism of Hydrocarbon Biosynthesis from Aldehyde in Selected Insect Species: Requirement for 0.2 and NADPH and Carbonyl Group Released as CO₂," *Insect Biochemistry Molecular Biology*, 1996, vol. 26, No. 2, pp. 203-208.
- Murata, N. et al., "Acyl-lipid Desaturases and Their Importance in the Tolerance and Acclimatization to Cold Cyanobacteria," *Biochem J.*, 1995, pp. 1-8, vol. 308.
- Murata, N. et al., "Modes of Fatty-Acid Desaturation in Cyanobacteria," *Plant Cell Physiology*, 1992, pp. 933-941, vol. 33, No. 7.
- PCT International Search Report and Written Opinion. PCT Application No. PCT/US2010/041619, Jan. 10, 2011, seventeen pages.
- PCT Invitation to Pay Additional Fees, etc., PCT Application No. PCT/US2010/041619, Sep. 30, 2010, two pages.
- Phung, L.T., "Genes for Fatty Acid Biosynthesis in the Cyanobacterium *Synechococcus*," Abstracts of the 951th General Meeting of the American Society for Microbiology, Jan. 1, 1995, p. 524.
- Phung, L.T. UniProt Direct Submission ID 054764, Nov. 1, 1996, p. 1 [Online] [Retrieved Dec. 7, 2010] Retrieved from the Internet <URL:<http://www.uniprot.org/uniprot/O54764.txt?version=1>>.
- Phung, L.T. UniProt Direct Submission ID 054765, Nov. 1, 1996, p. 1 [Online] [Retrieved from the Internet <URL:<http://www.uniprot.org/uniprot/Q54765.txt?version=1>>.
- Qi, Q. et al., "Application of the *Synechococcus nirA* Promoter to Establish an Inducible Expression System for Engineering the *Synechocystis* Tocopherol Pathway," *Applied and Environmental Microbiology*, Oct. 2005, pp. 5678-5684, vol. 71, No. 10.
- Reed, J.R. et al., Proposed Mechanism for the Cytochrome P450-Catalyzed Conversion of Aldehydes to Hydrocarbons in the House Fly, *Musca domestica*, *Biochemistry*, 1995, vol. 34, pp. 16221-16227.
- Reed, J.R. et al., "Unusual Mechanism of Hydrocarbon Formation in the Housefly: Cytochrome P450 Converts Aldehyde to the Sex Pheromone Component (Z)-9-tricosene and CO₂," *Proceedings of the National Academy of Science USA*, Oct. 1994, pp. 1000-10004, vol. 91.
- Rude, M.A. et al., "New Microbial Fuels: A Biotech Perspective," *Current Opinion in Microbiology*, 2009, vol. 12, No. 3, pp. 274-281.
- Schirmer, A. et al., "Microbial Biosynthesis of Alkanes," *Science*, Jul. 30, 2010, pp. 559-562, vol. 329.
- State Intellectual Property Office of the People's Republic of China, Office Action, Chinese Patent Application No. 201080031073.7. Jul. 8, 2013, twelve pages.
- Sugita, C. et al., "Complete Nucleotide Sequence of the Freshwater Unicellular Cyanobacterium *Synechococcus elongates* PCC 6301 Chromosome. Gene Content and Organization," *Photosynthesis Research: Official Journal of the International Society of Photosynthesis Research*, 2007, pp. 55-67, vol. 93.
- Welsh, E.A. et al. UniProt Direct Submission B1WR71, May 20, 2008, p. 1 [Online] [Retrieved Dec. 7, 2010] Retrieved from the Internet: <URL:<http://www.uniprot.org/uniprot/B1WR71.txt?version=1>>.
- Welsh, E.A. et al. UniProt Direct Submission B1WWQ5, May 20, 2008, p. 1 [Online] [Retrieved Dec. 7, 2010] Retrieved from the Internet <URL:<http://www.uniprot.org/uniprot/B1WWQ5.txt?version=1>>.
- Wentzel, A. et al., "Bacterial Metabolism of Long-chain n-Alkanes," *Applied Microbiology and Biotechnology*, 2007, vol. 76, pp. 1209-1221.
- Winters, K. et al., "Hydrocarbons of Blue-Green Algae: Geochemical Significance," *Science*, 1969, pp. 467-468, vol. 163, No. 3866.
- United States Office Action, U.S. Appl. No. 13/051,807, Sep. 7, 2011, eight pages.
- United States Office Action for U.S. Appl. No. 13/913,795, Feb. 20, 2015, 12 pages.
- Office Action for Canadian Application No. 2,766,204, Dec. 22, 2014, 5 pages.
- Patent Examination Report No. 1 for Australian Application No. AU 2011302092, Mar. 17, 2015, 3 pages.

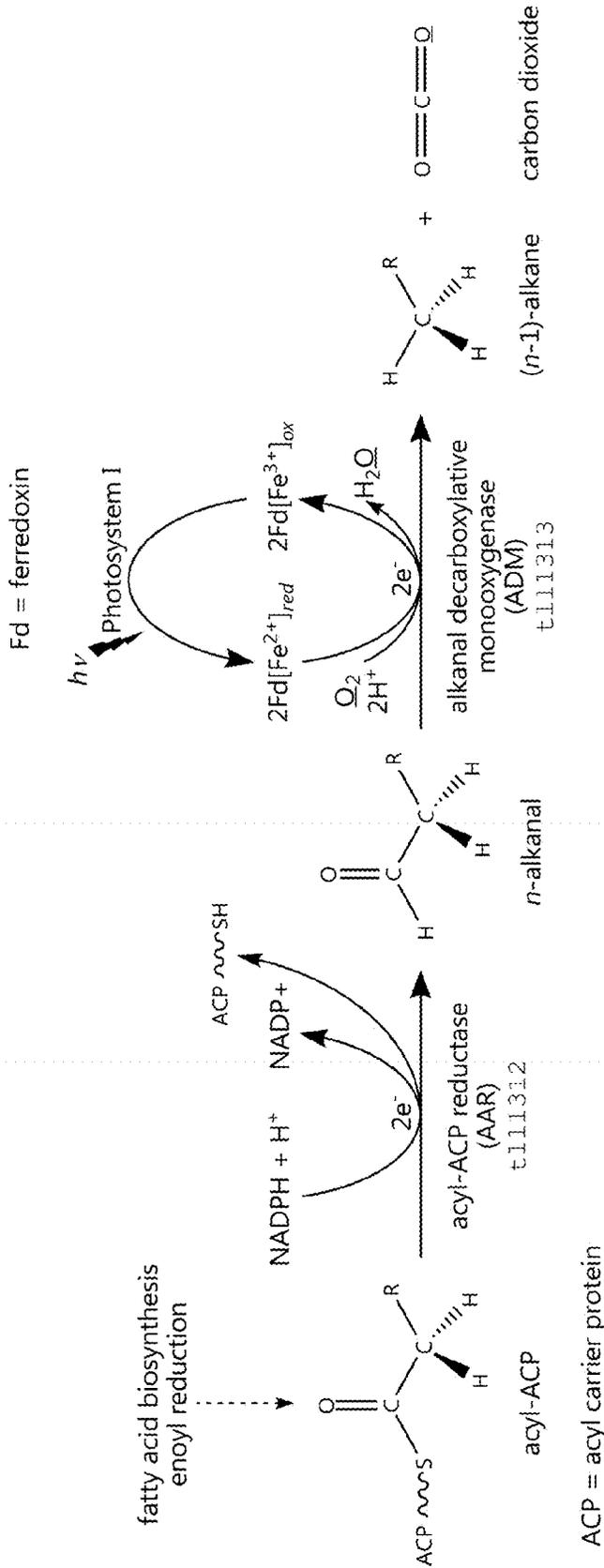


FIG. 1A

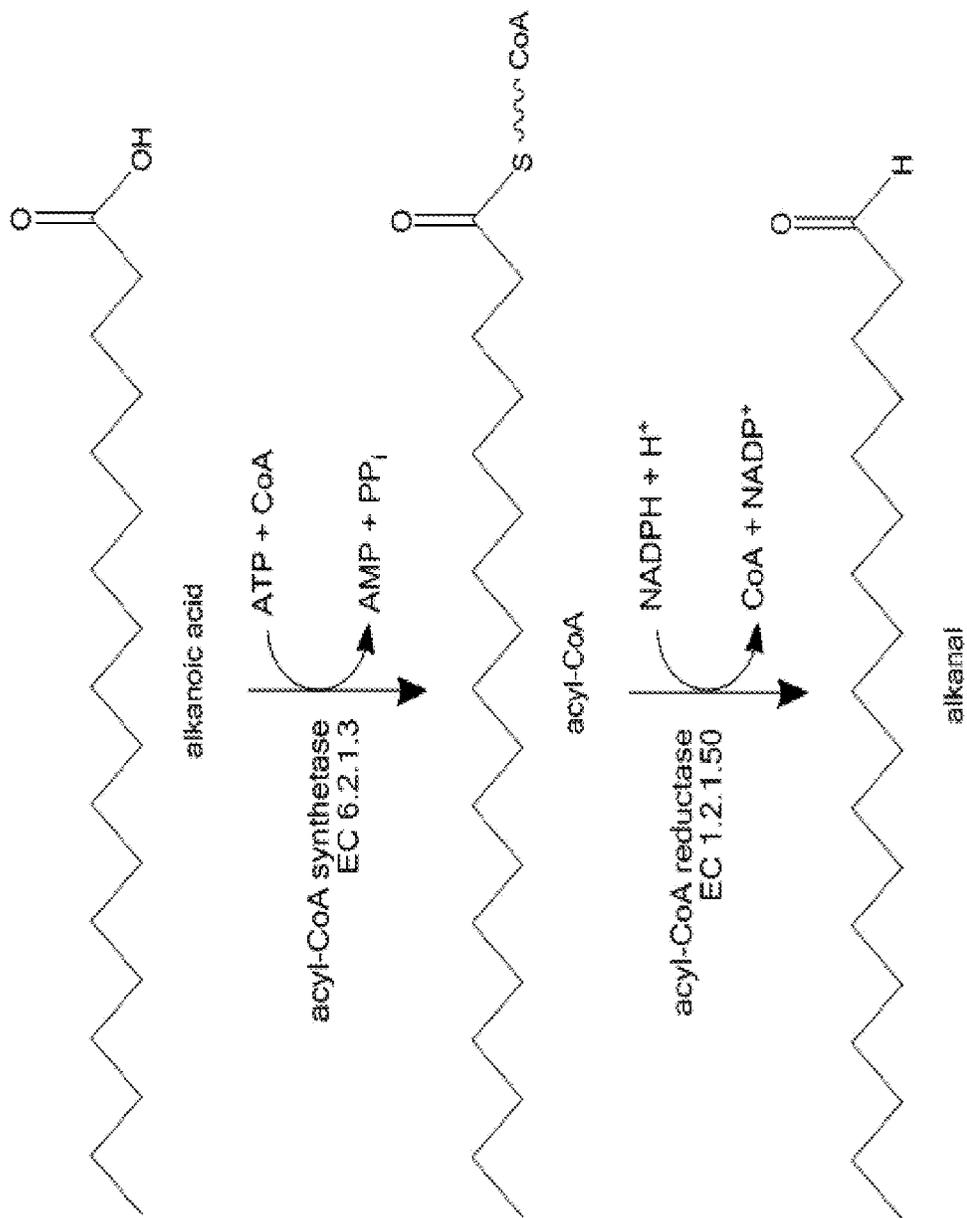


FIG. 1B

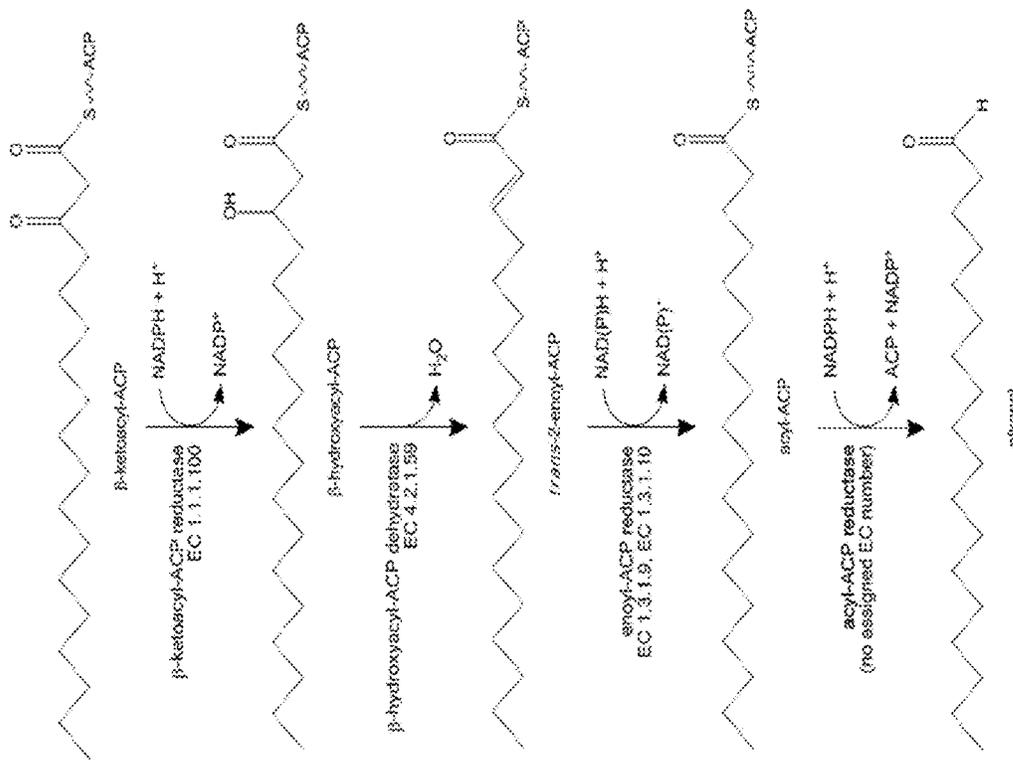


FIG. 1C

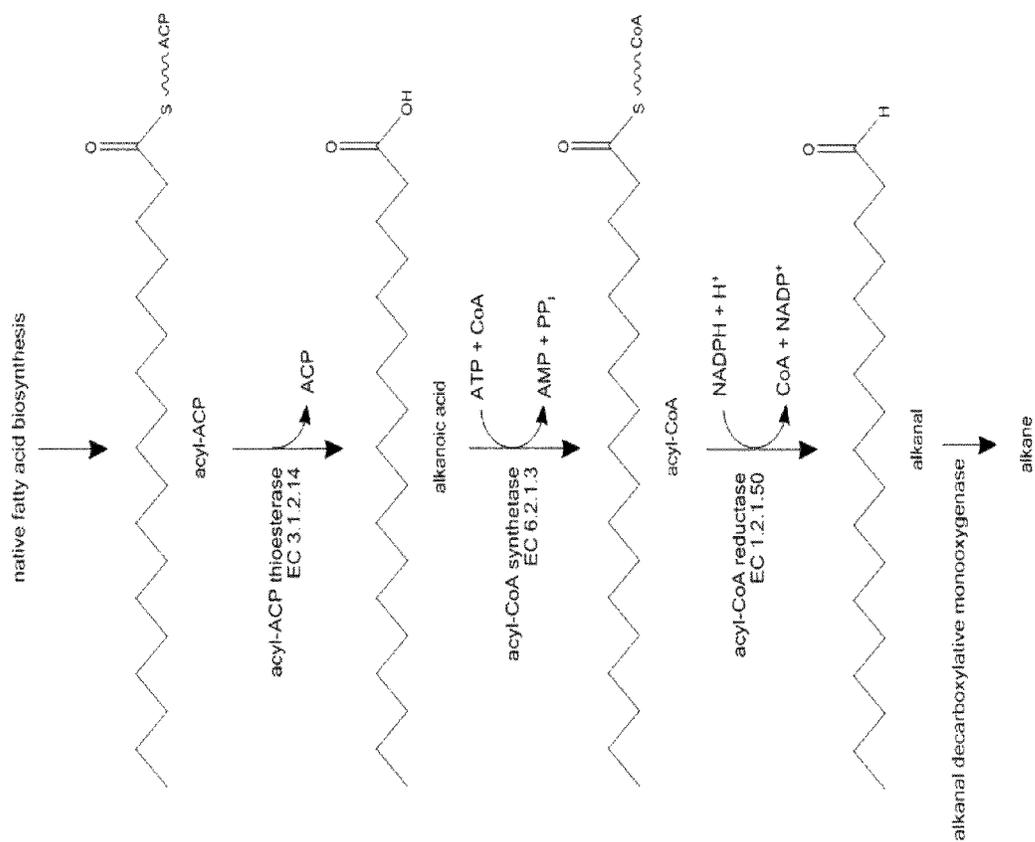


FIG. 1D

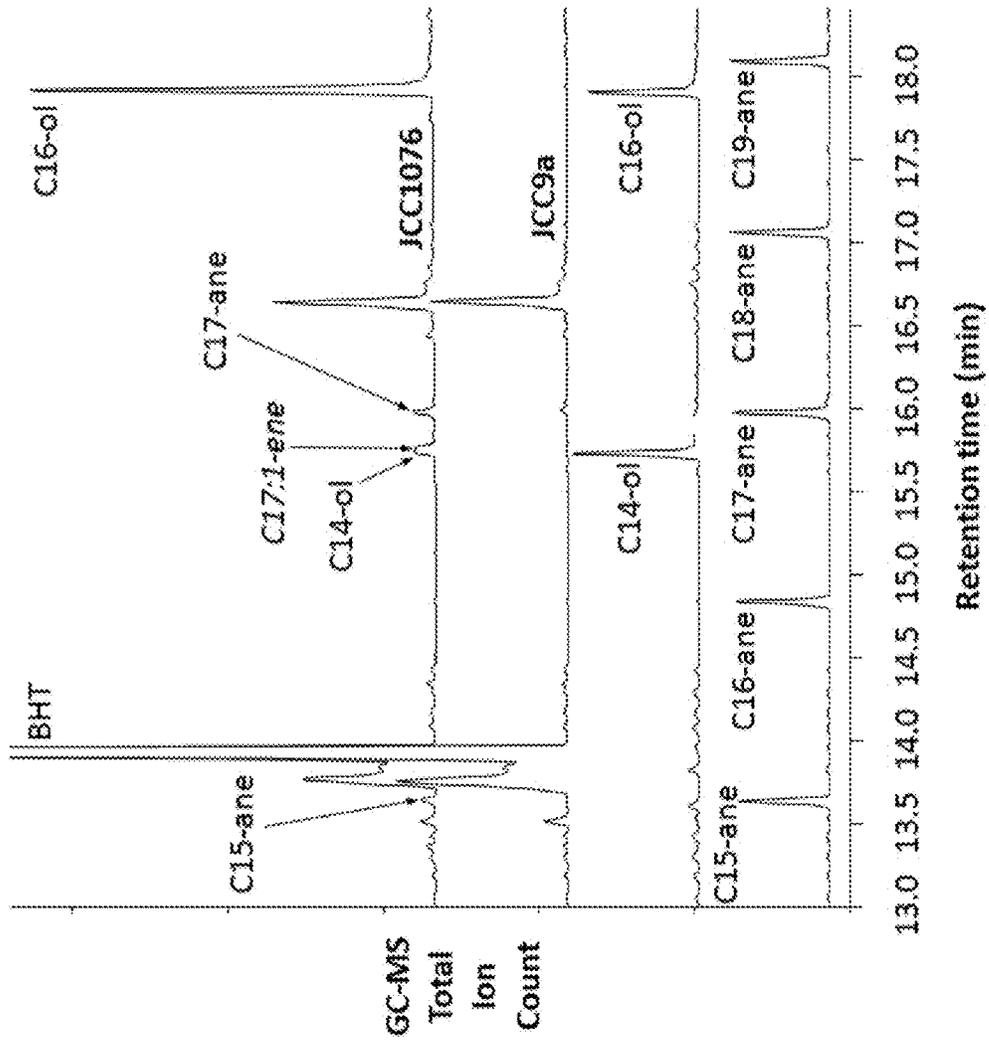


FIG. 2

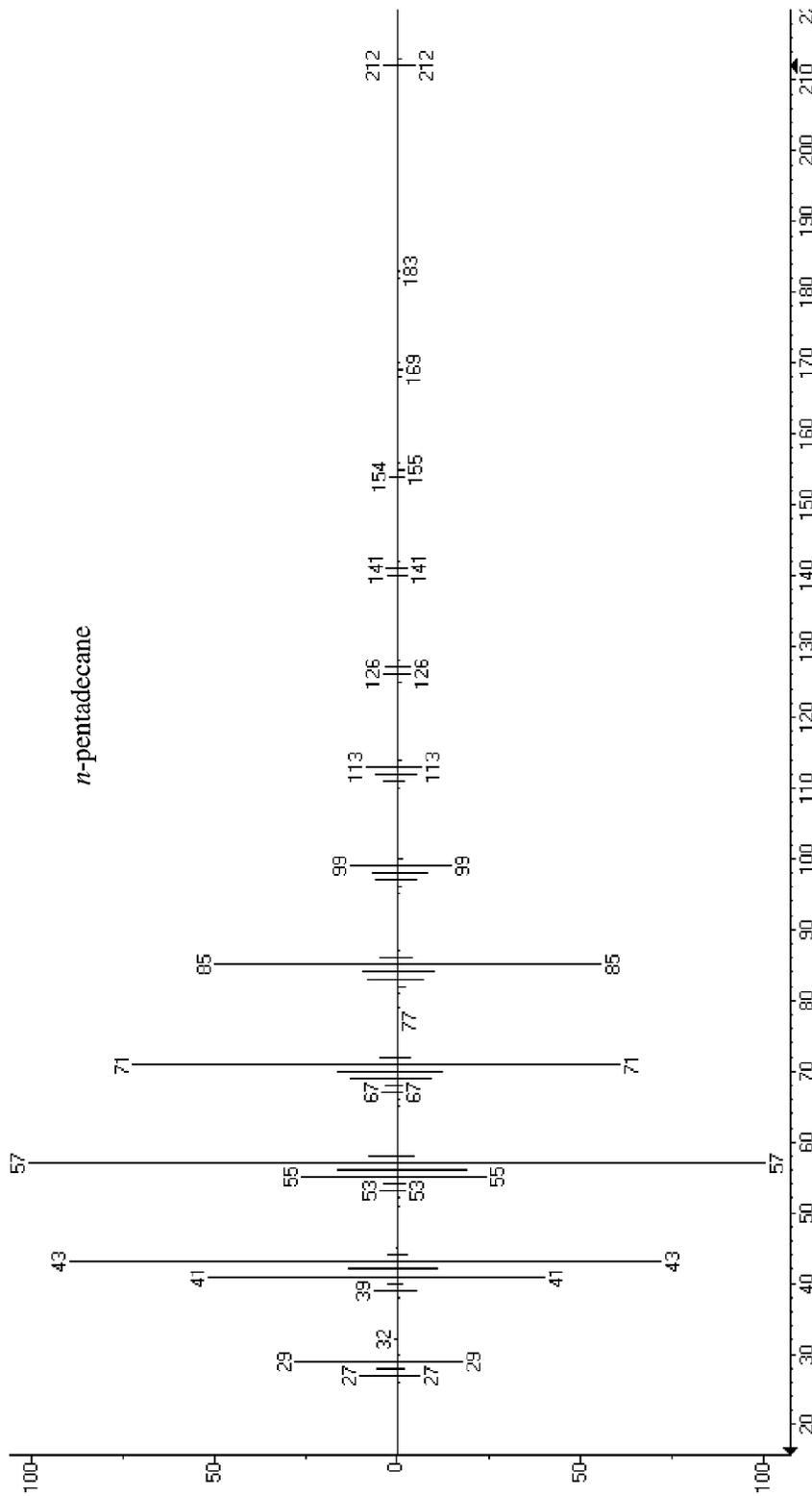


FIG. 3A

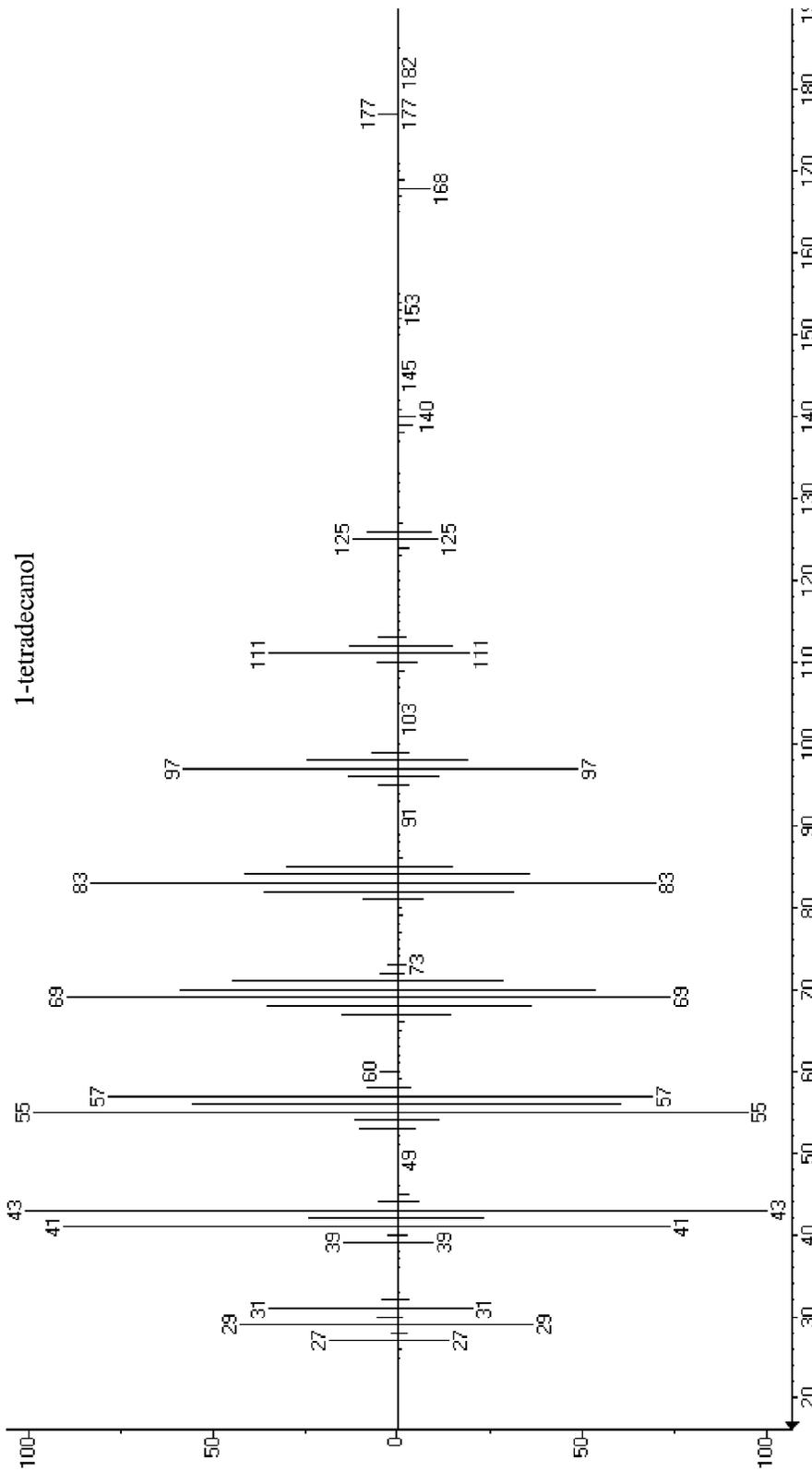


FIG. 3B

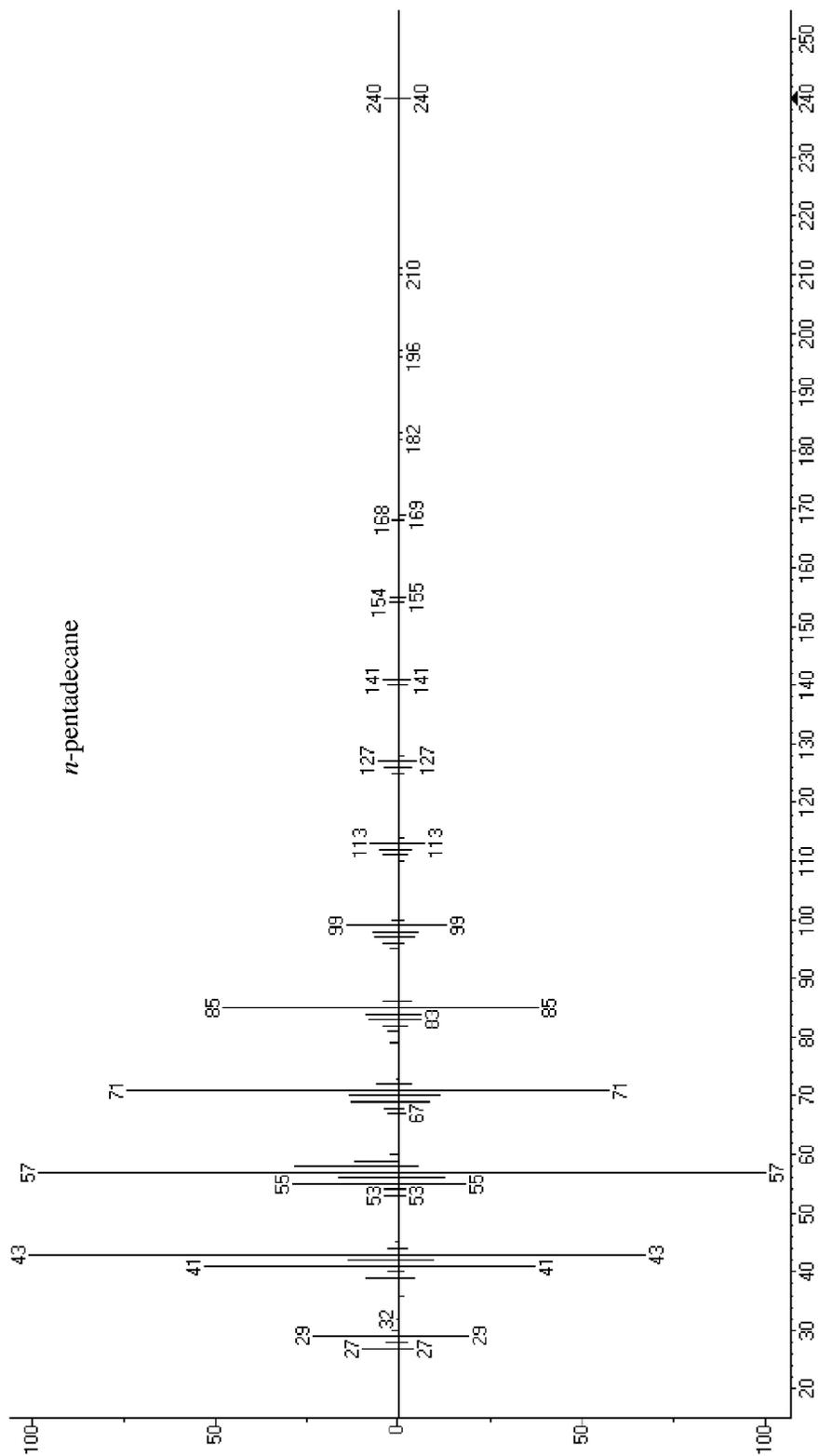


FIG. 3C

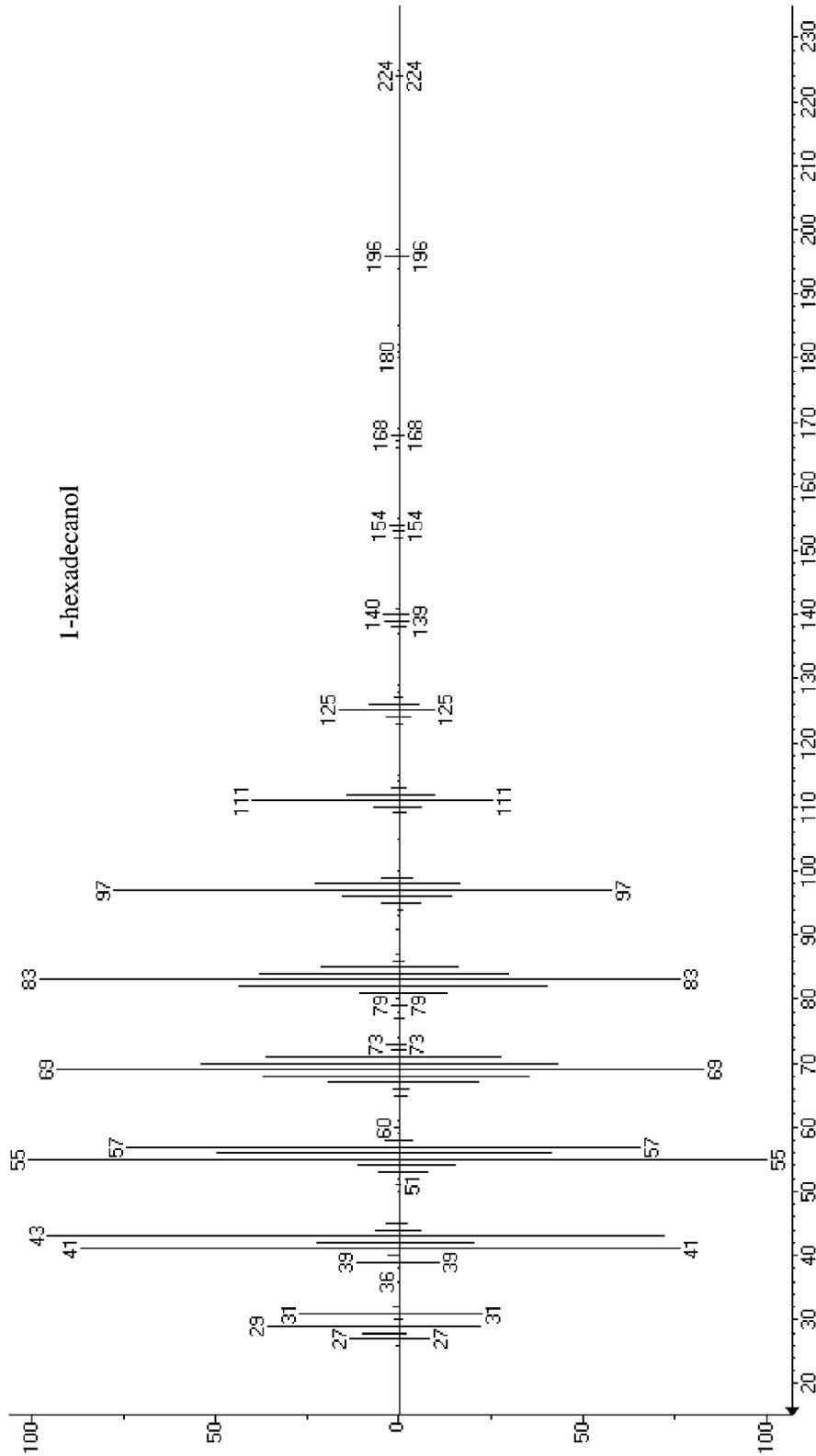


FIG. 3D

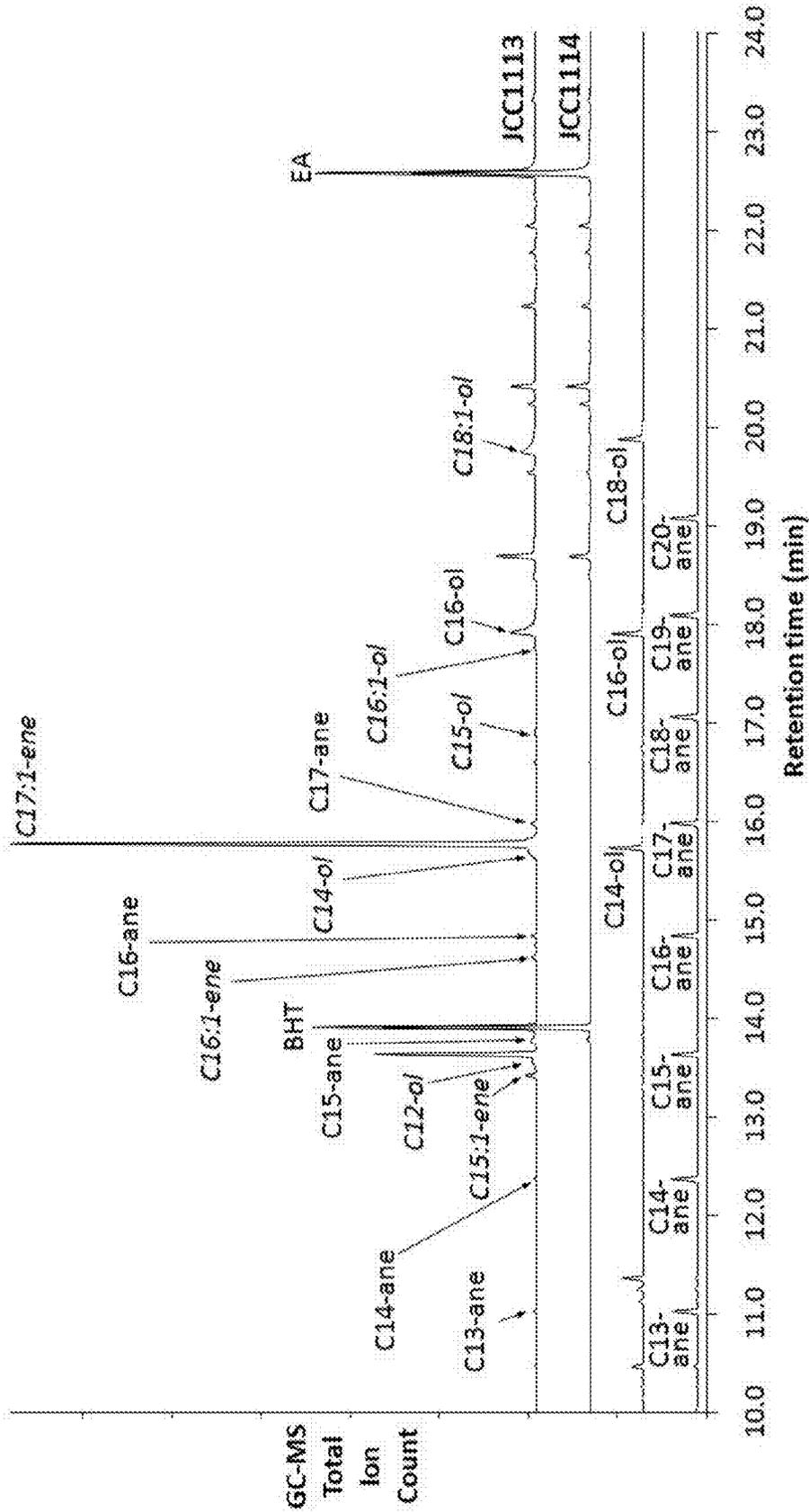


FIG. 4A

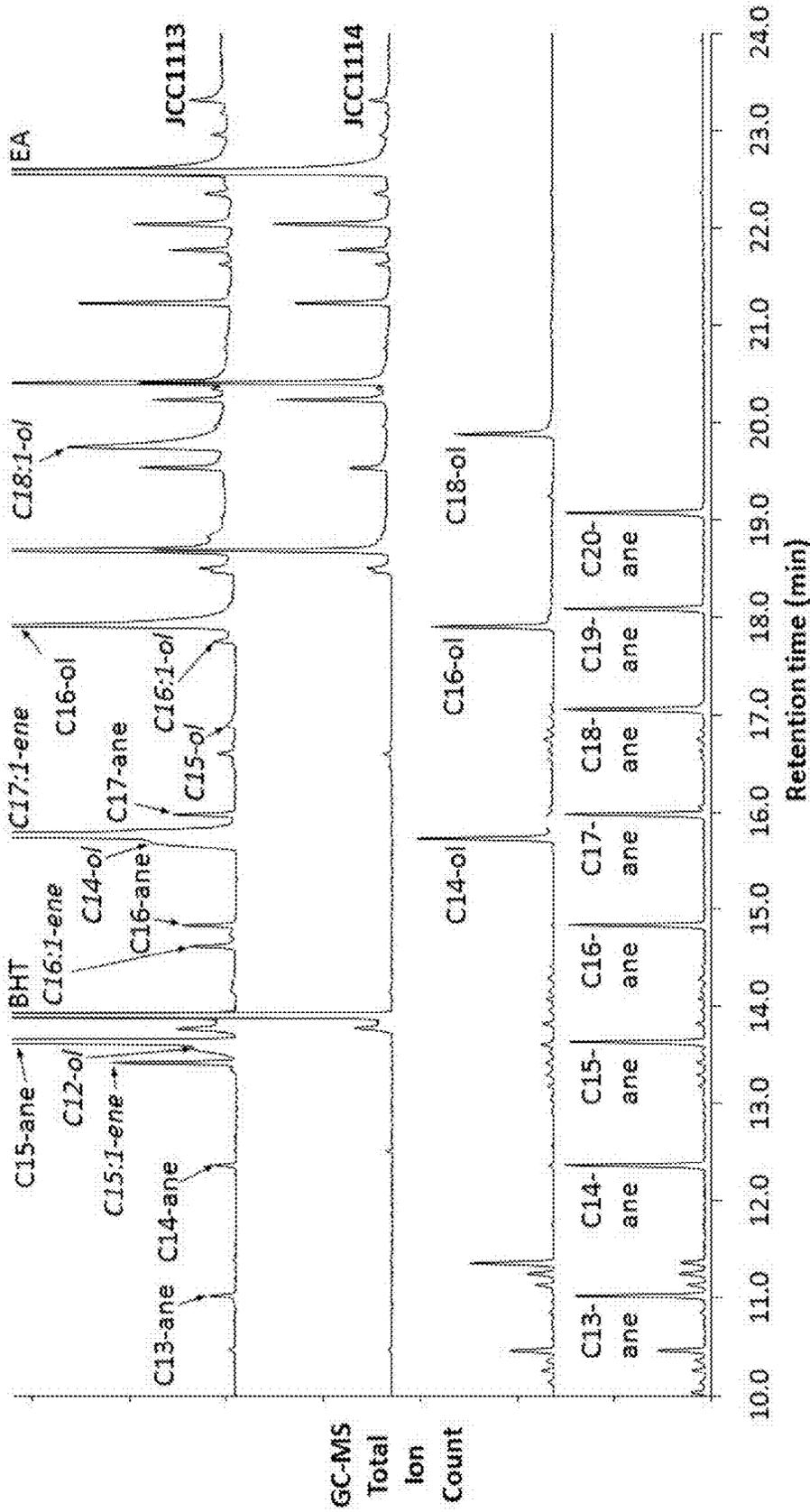


FIG. 4B

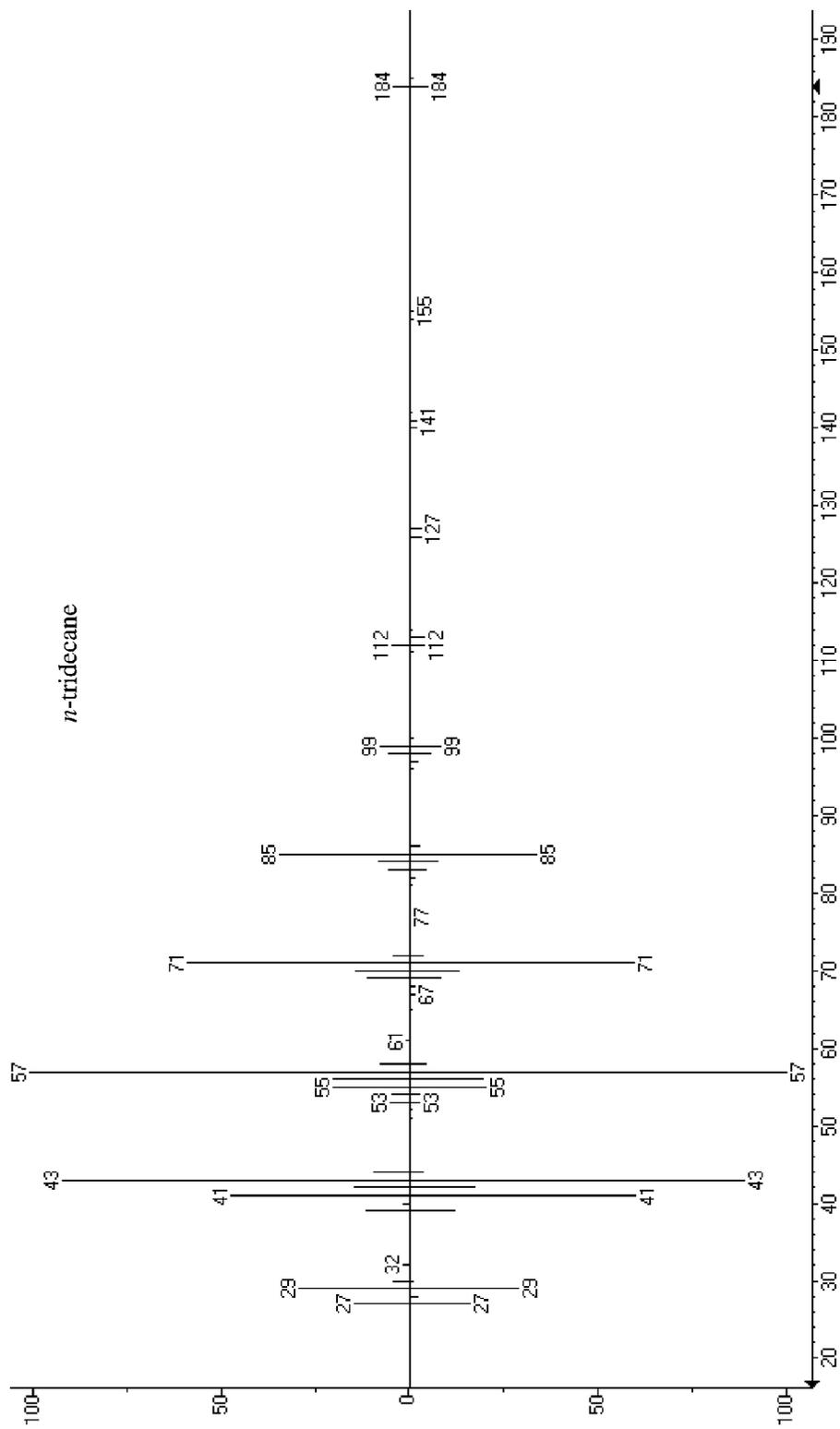


FIG. 5A

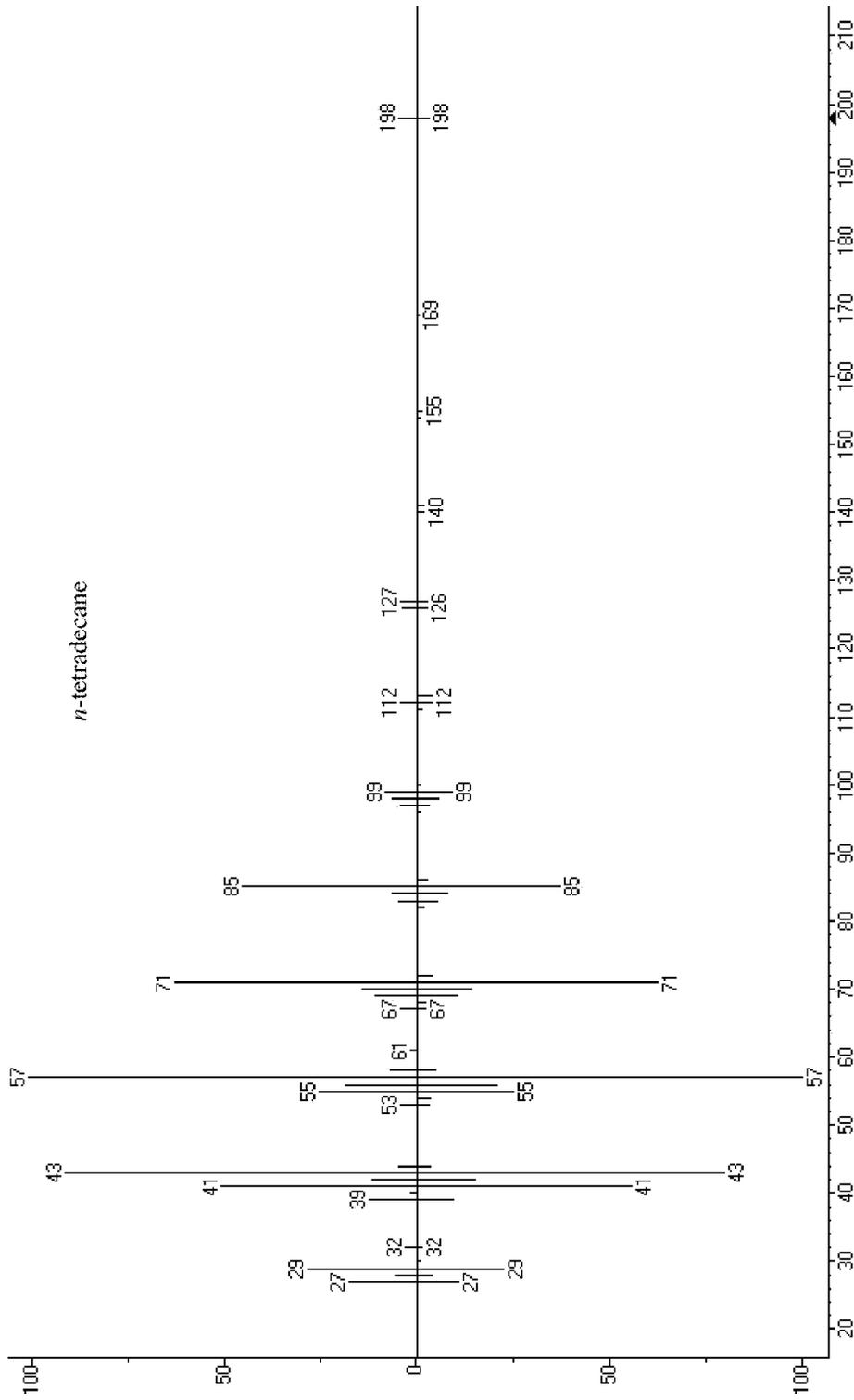


FIG. 5B

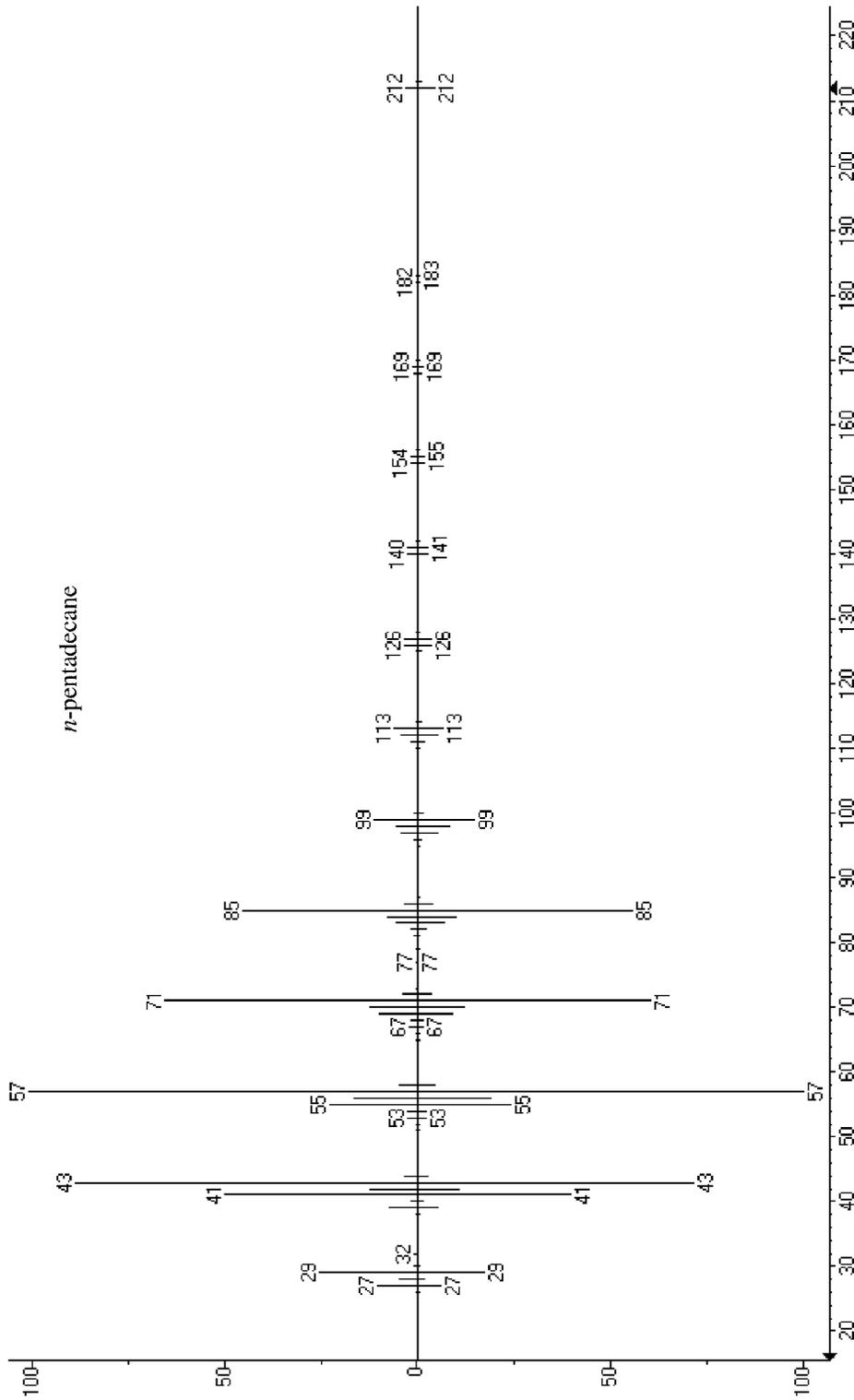


FIG. 5C

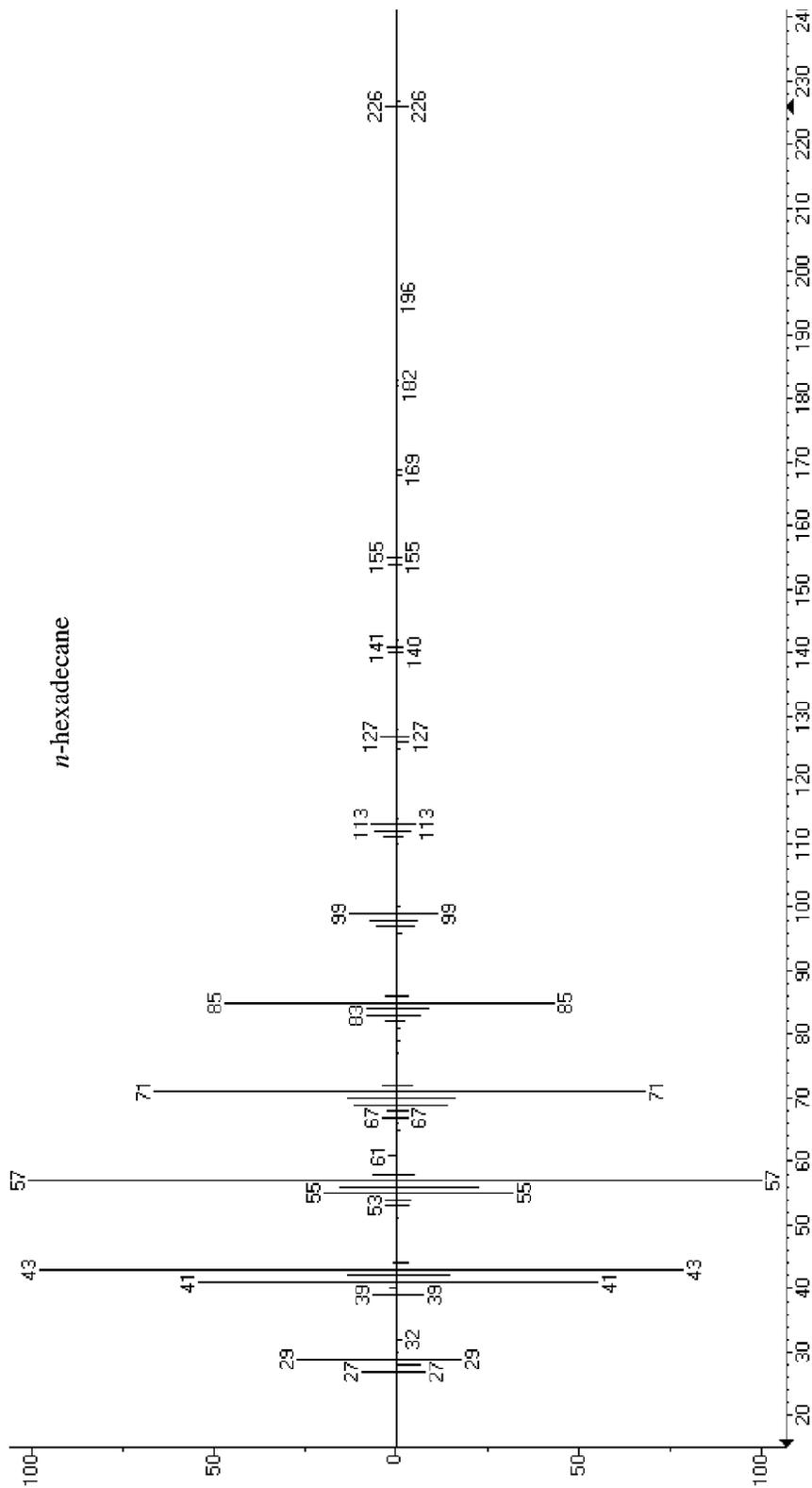


FIG. 5D

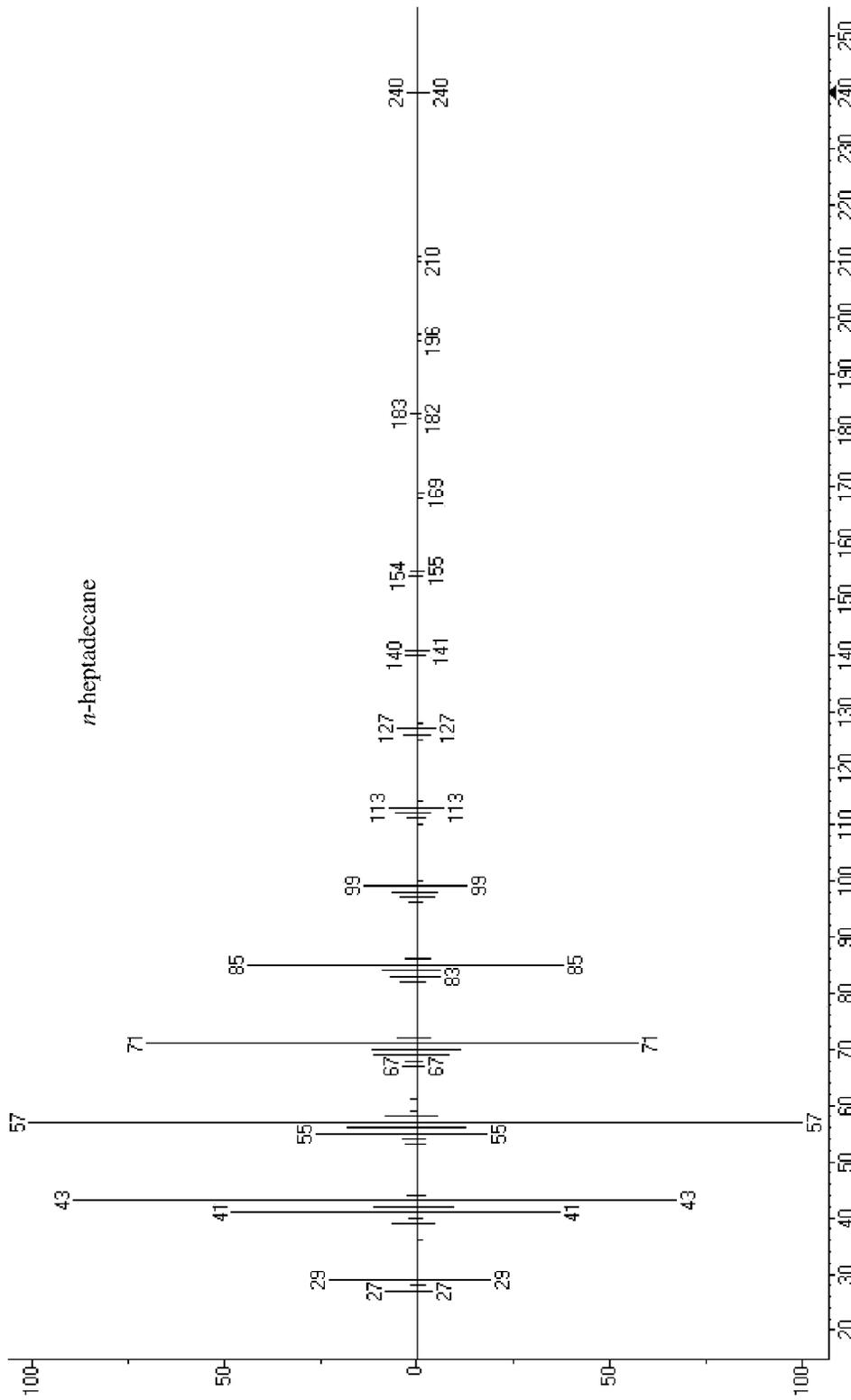


FIG. 5E

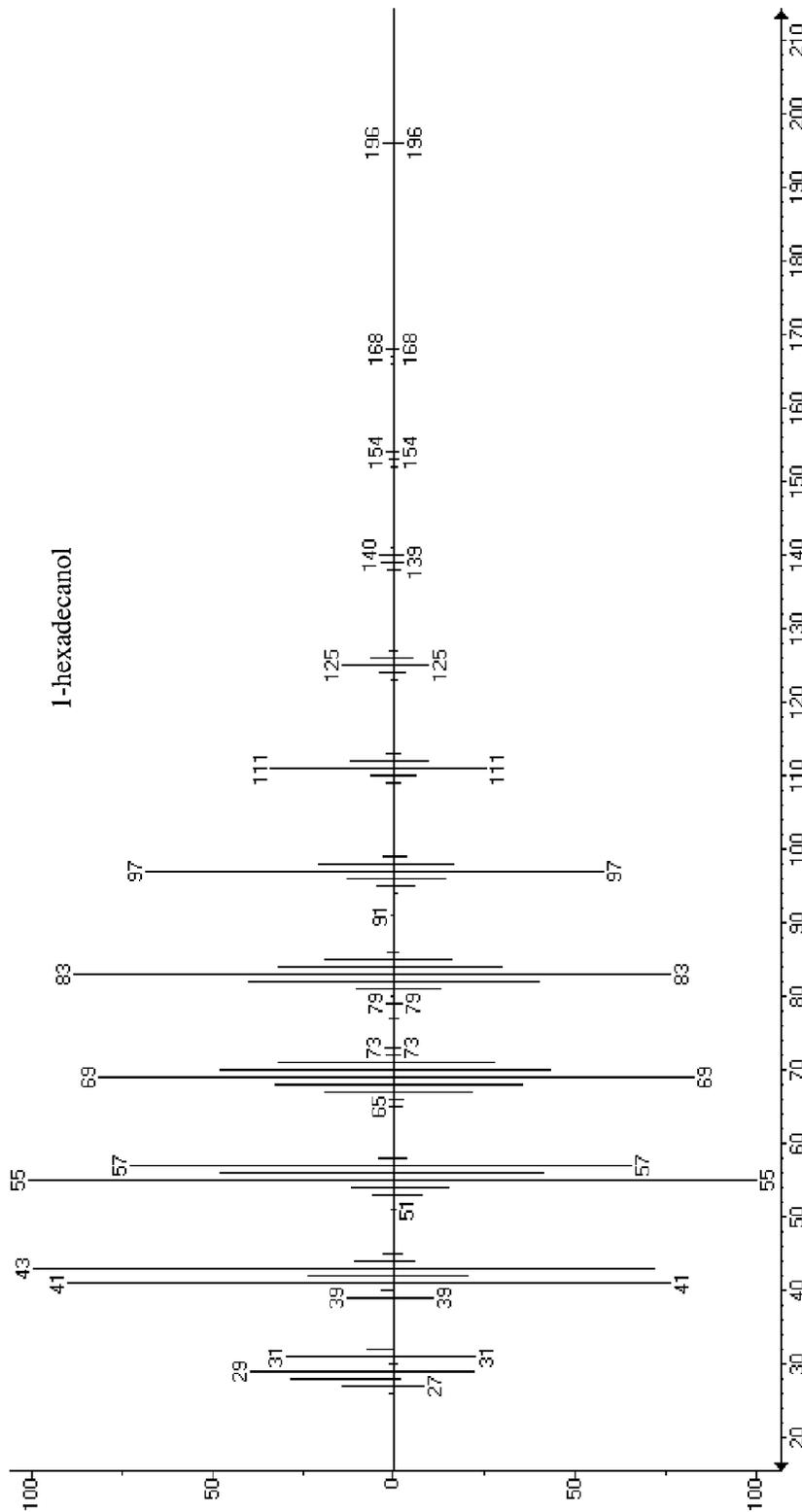


FIG. 5F

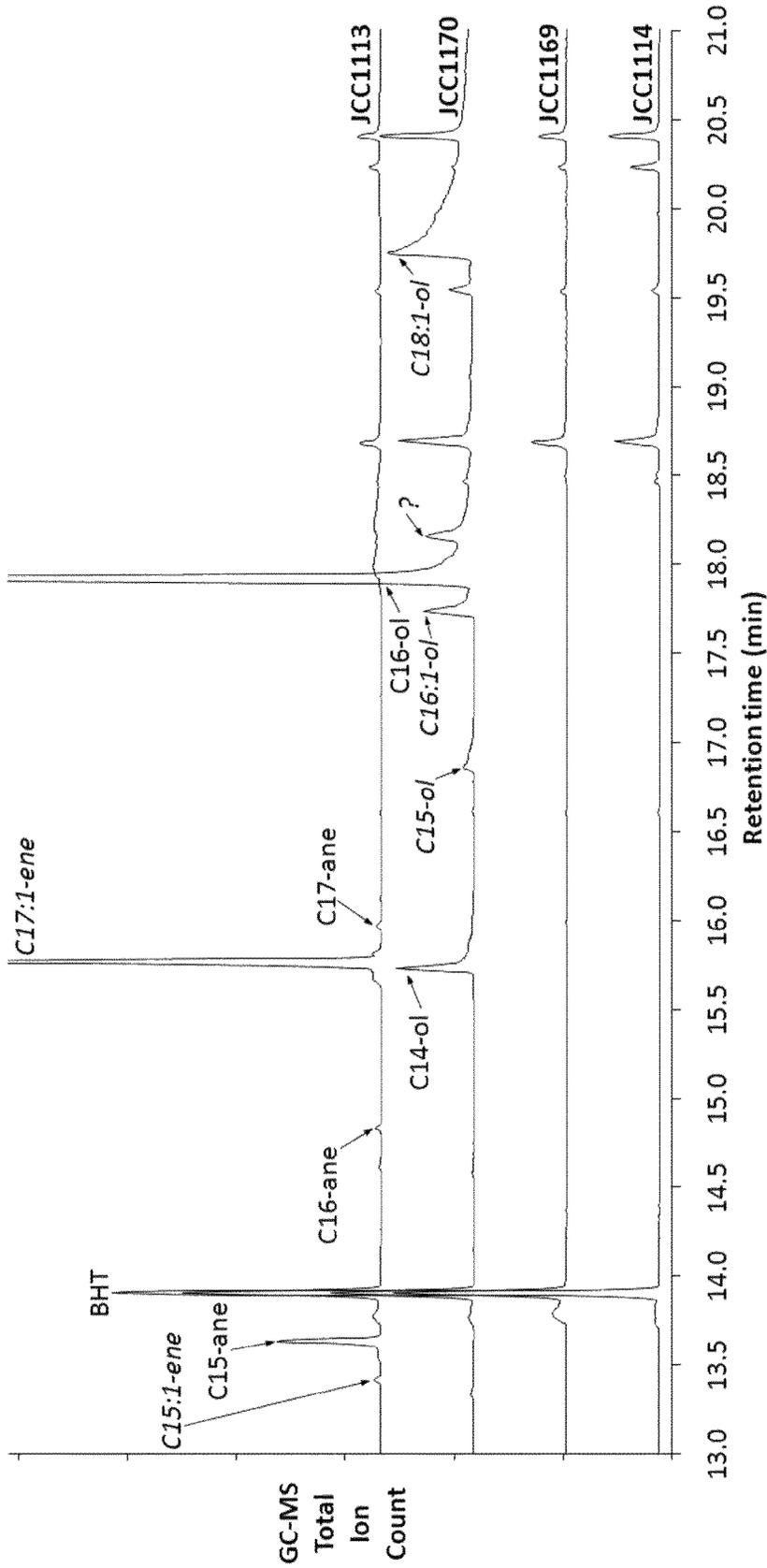


FIG. 6

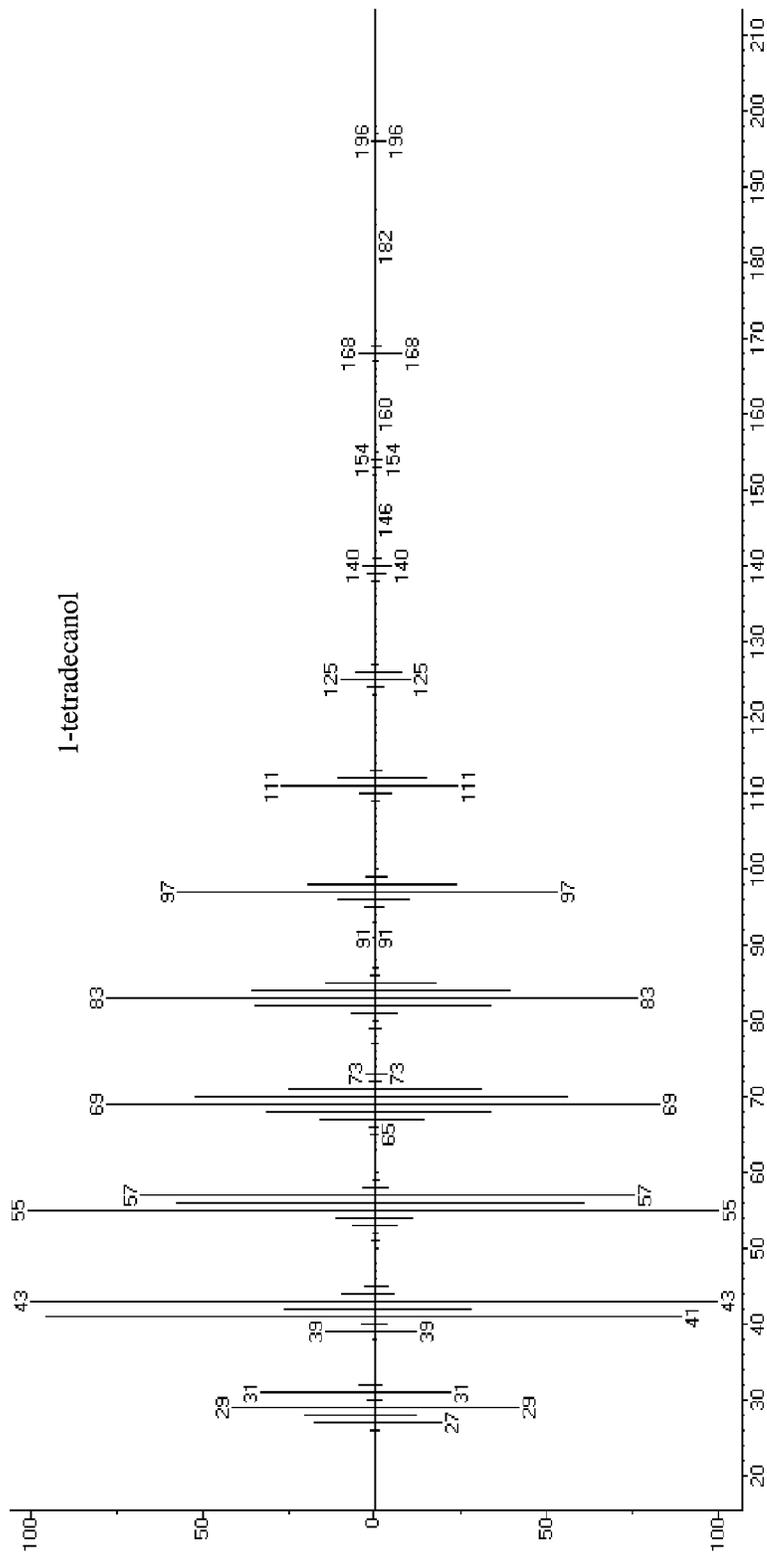


FIG. 7A

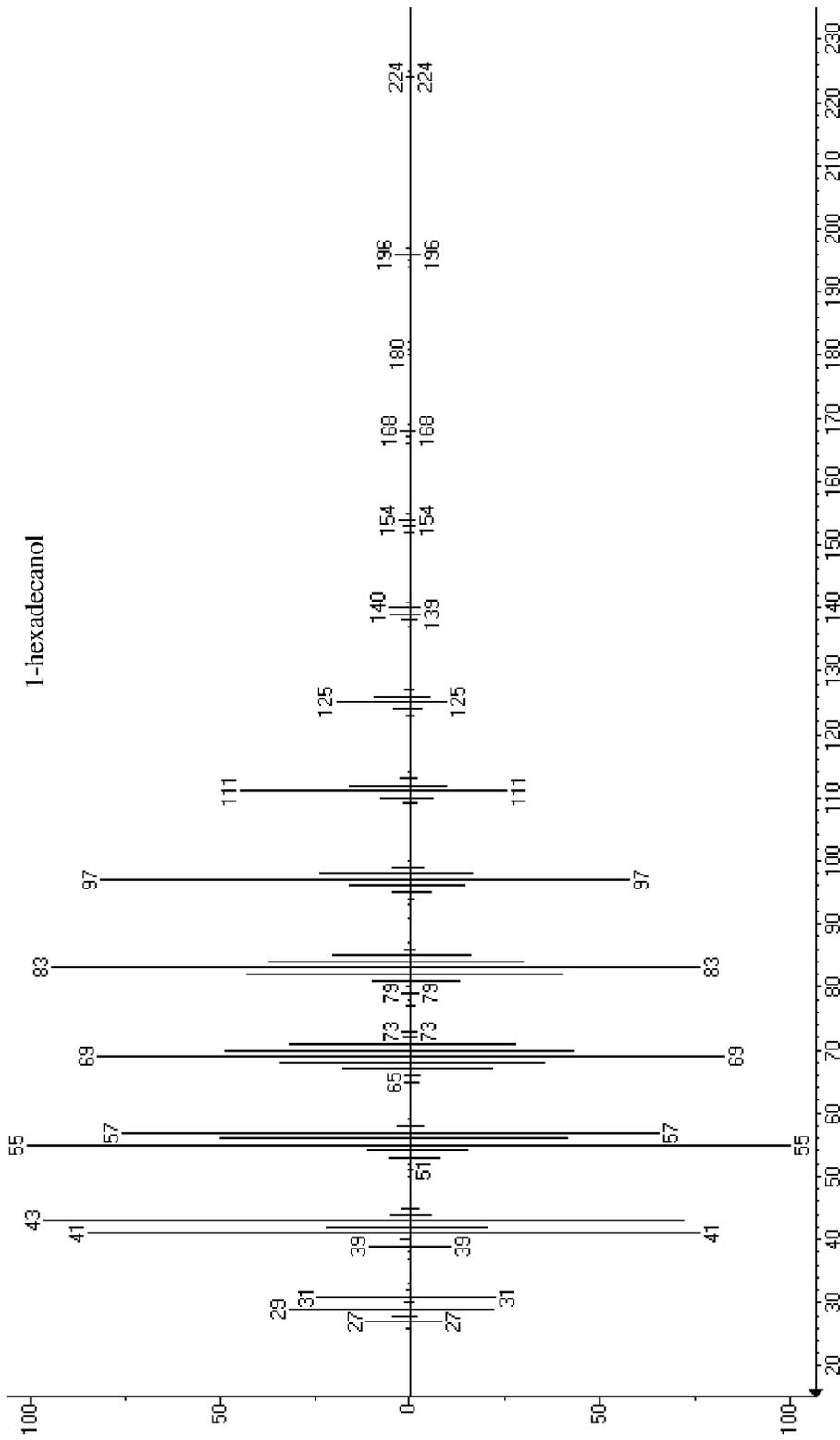


FIG. 7B

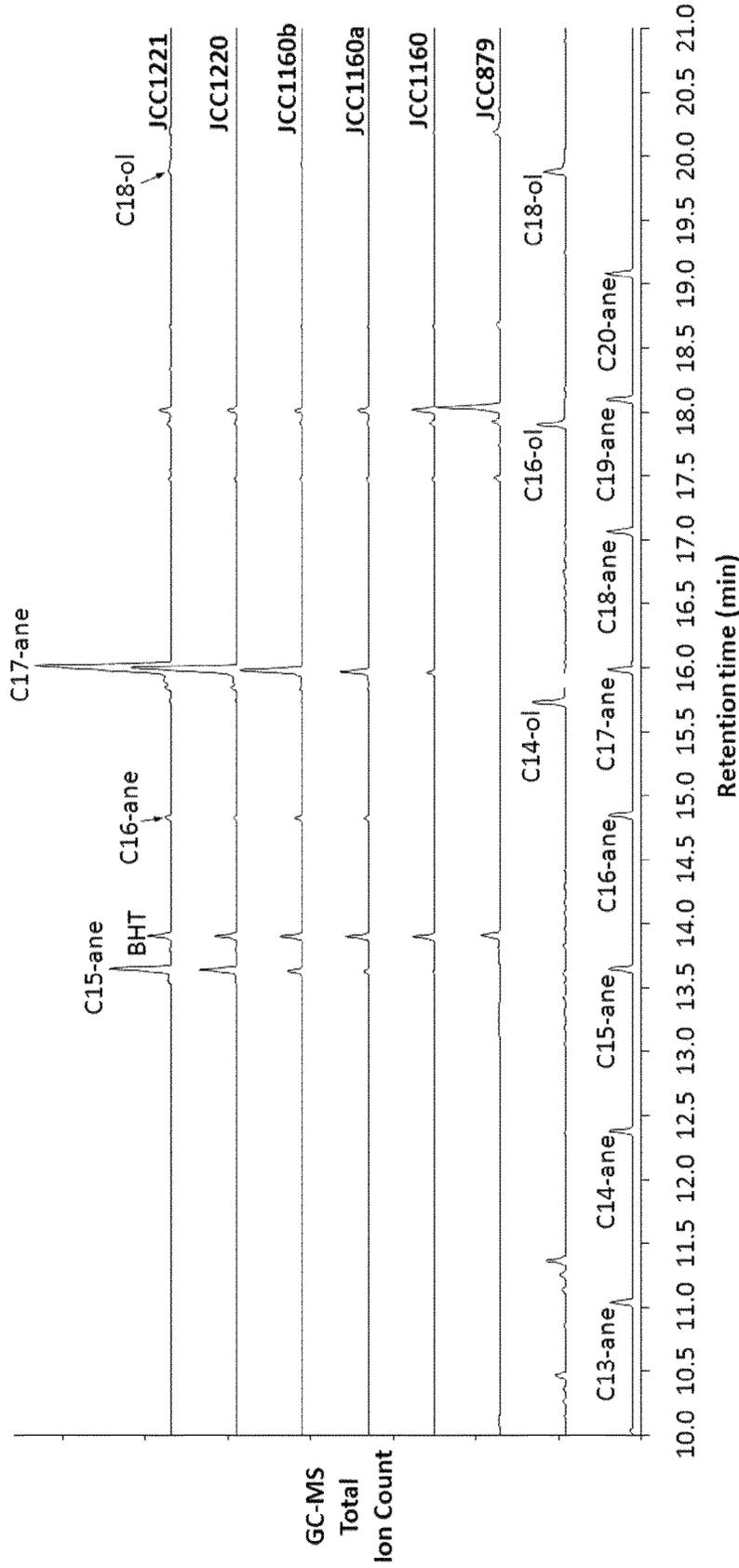


FIG. 8A

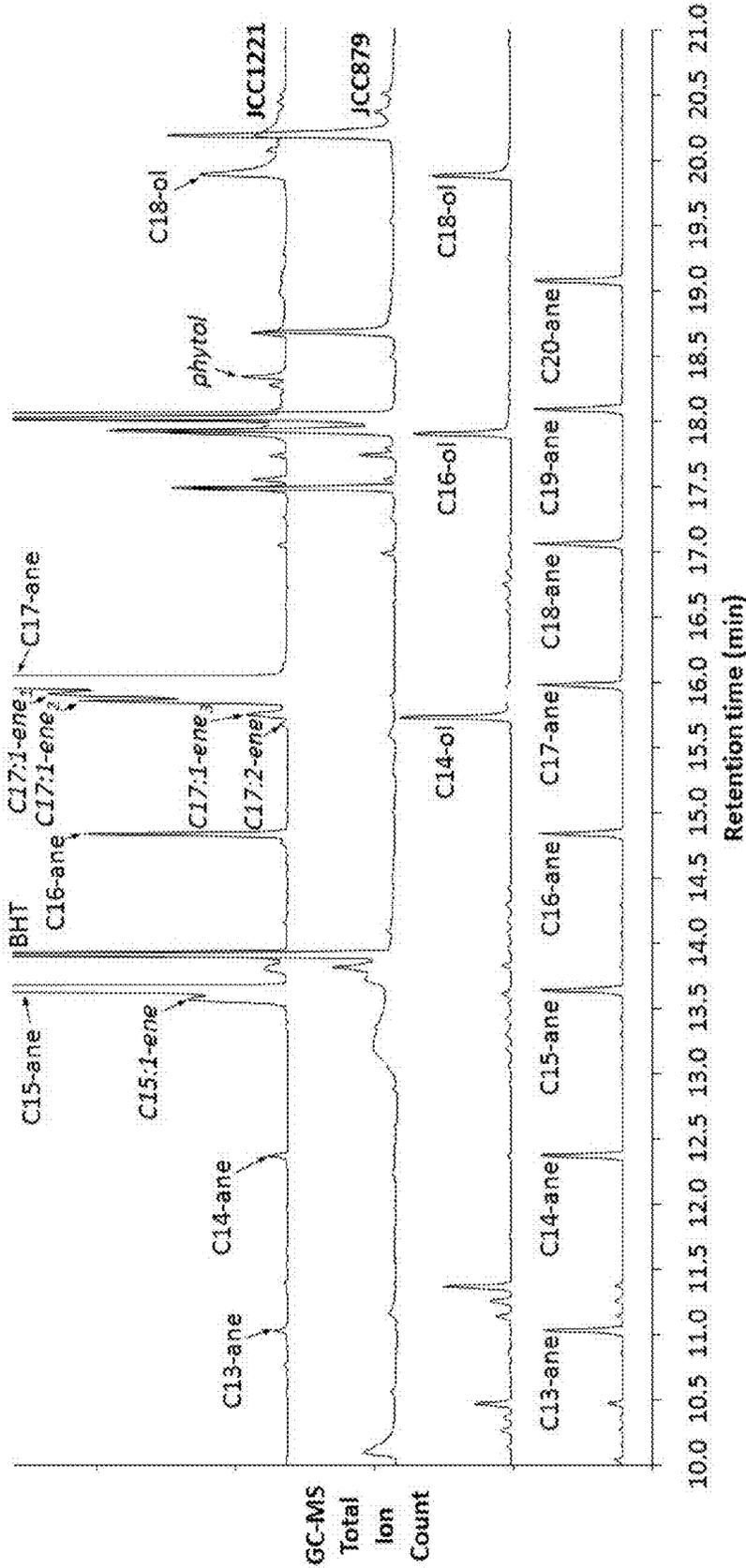


FIG. 8B

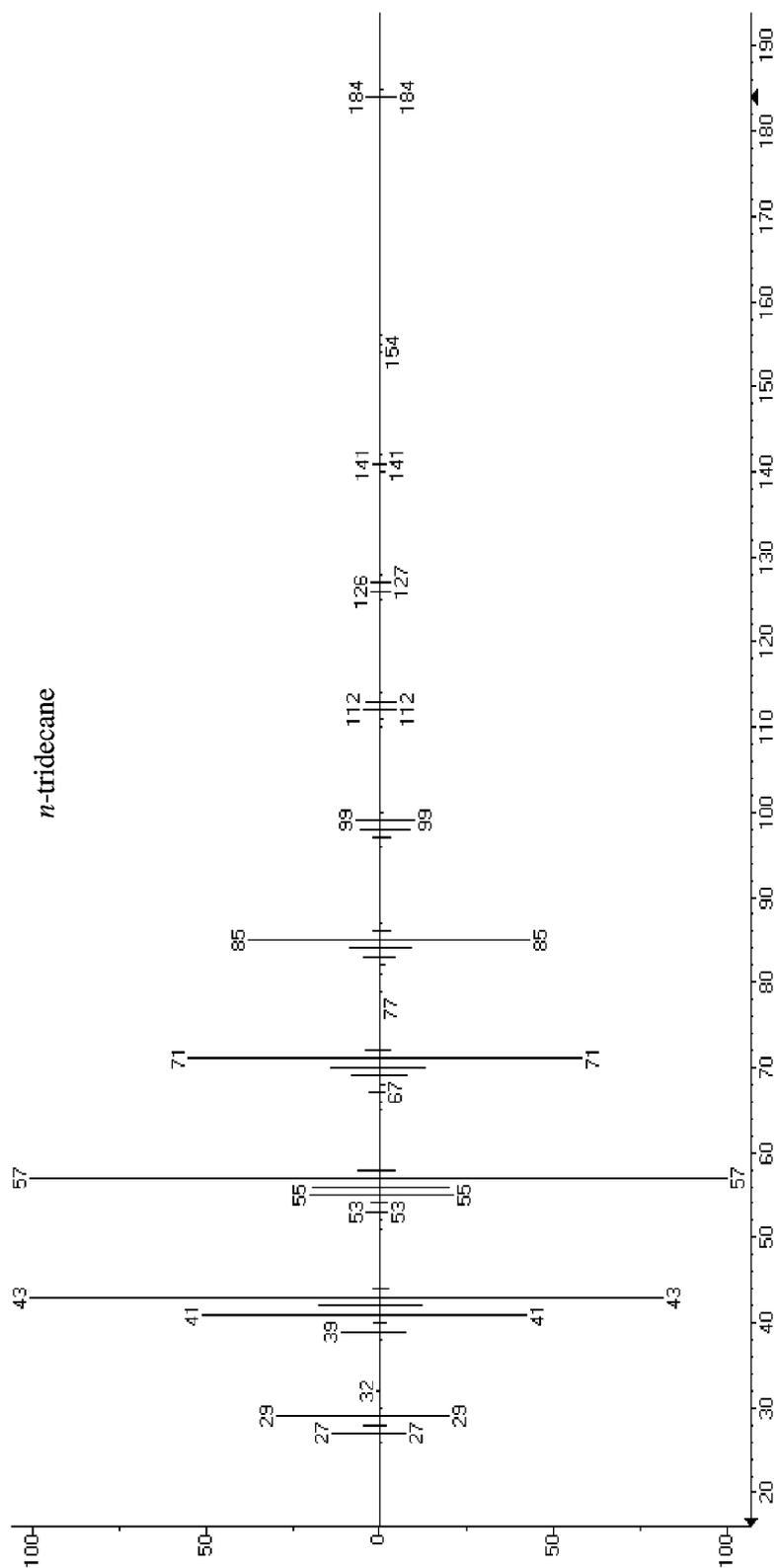


FIG. 9A

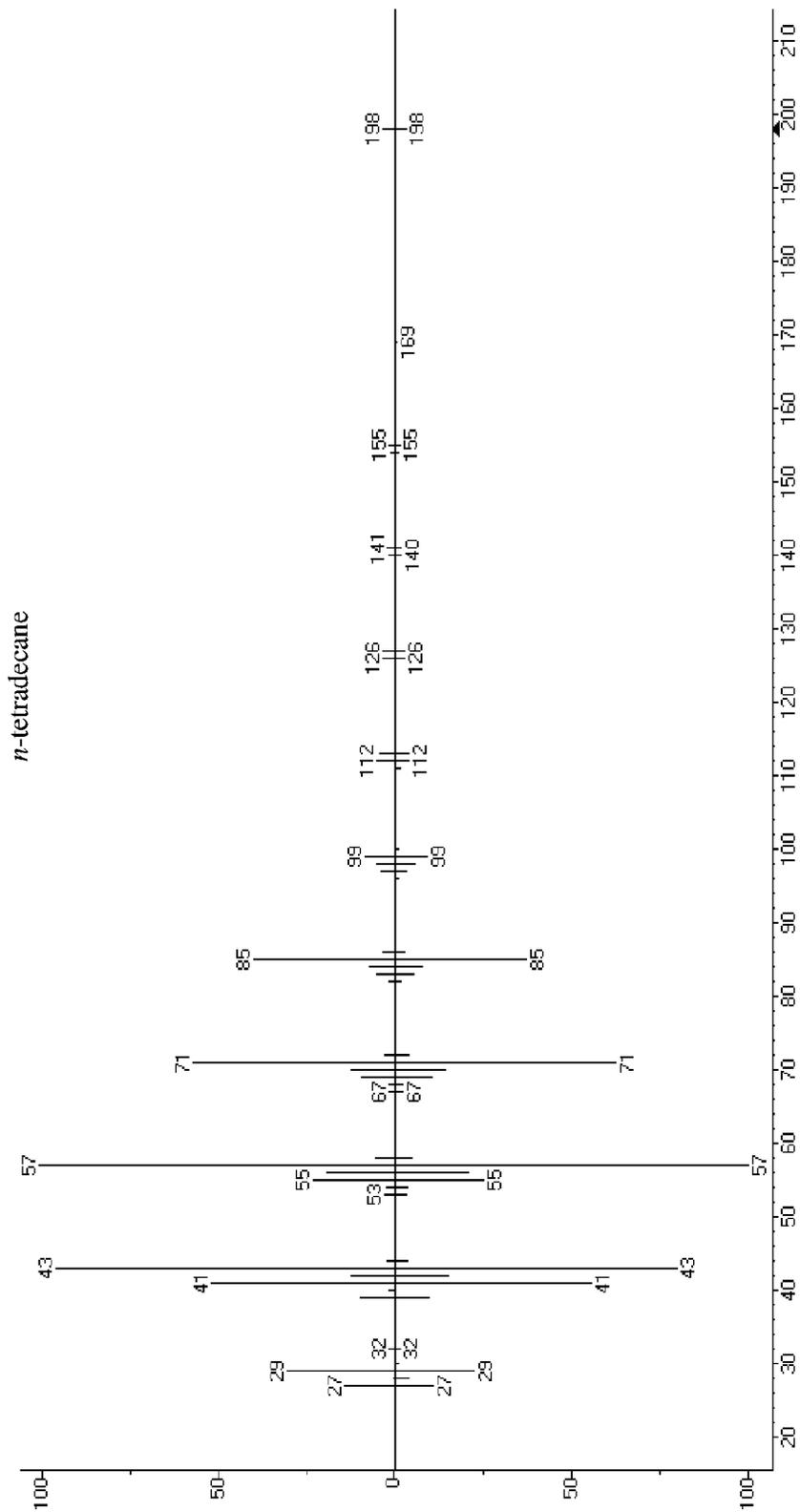


FIG. 9B

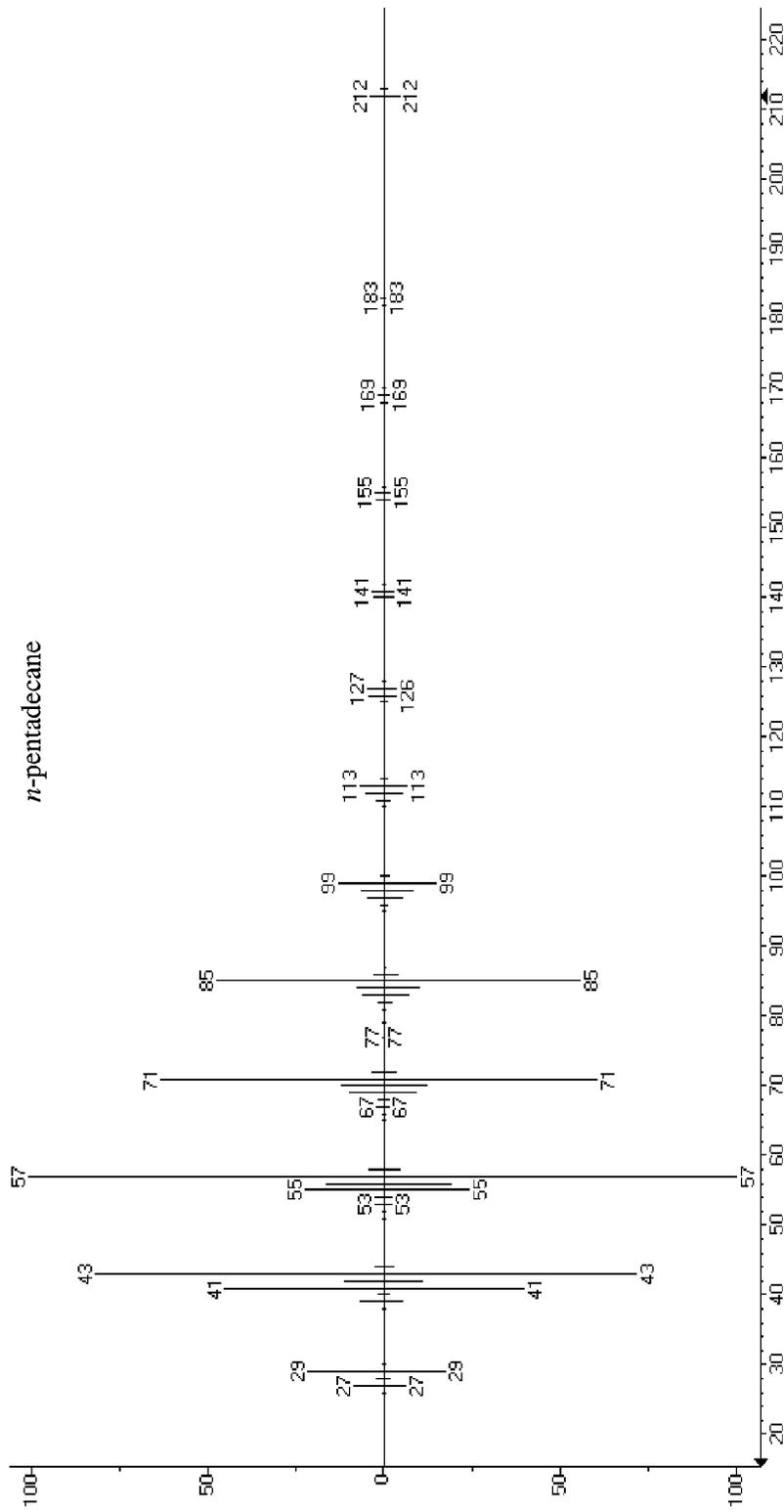


FIG. 9C

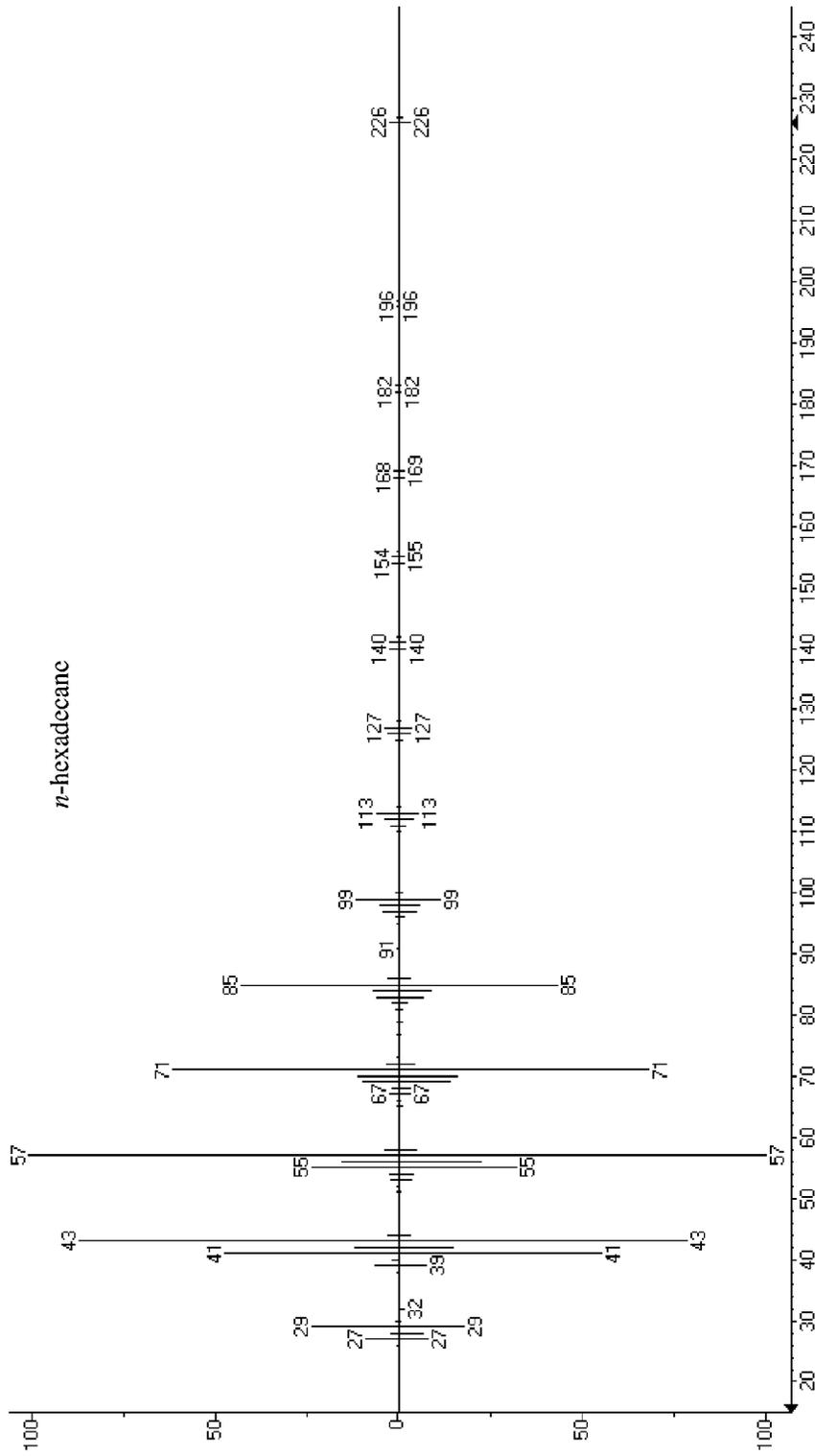


FIG. 9D

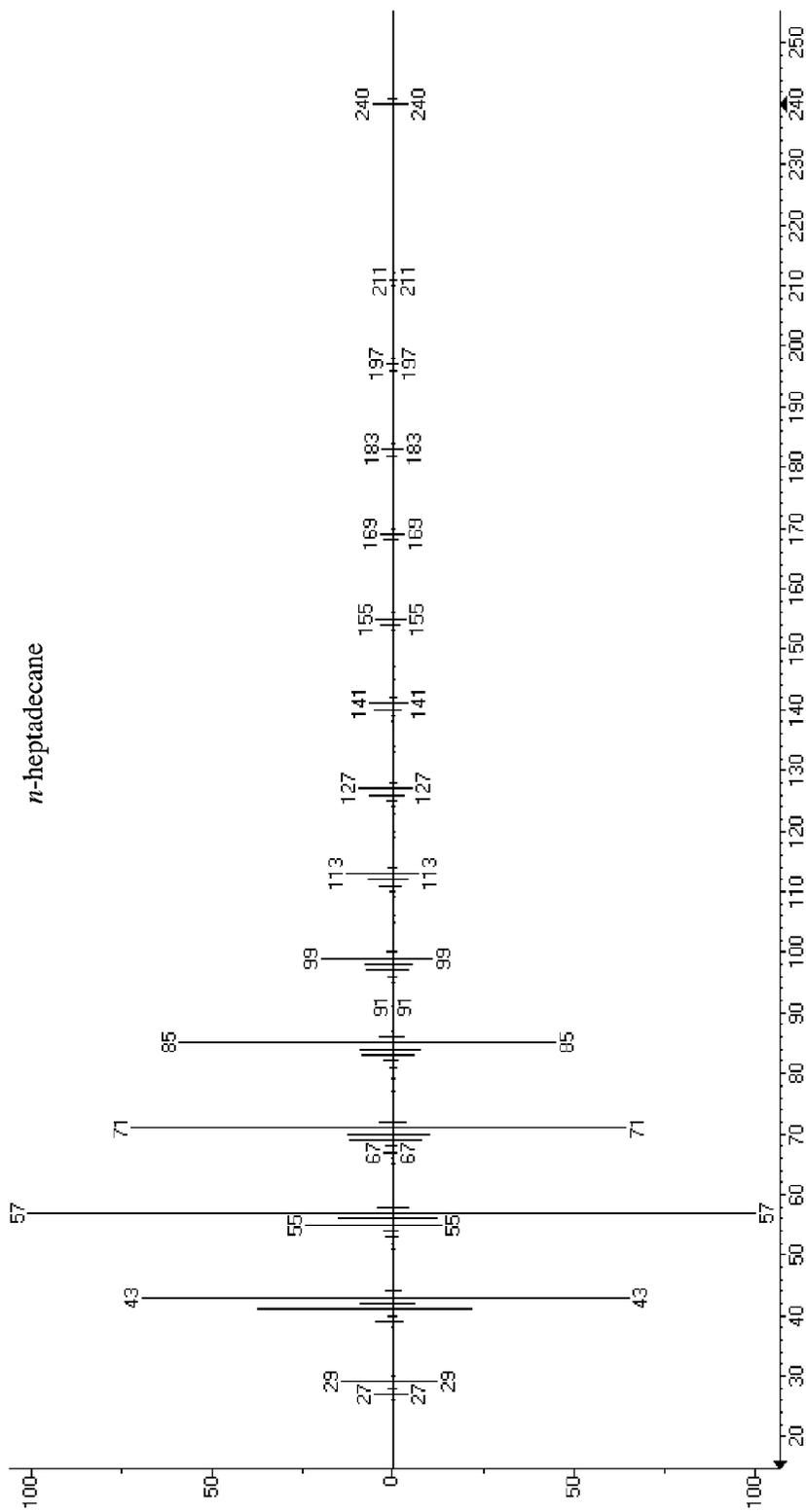


FIG. 9E

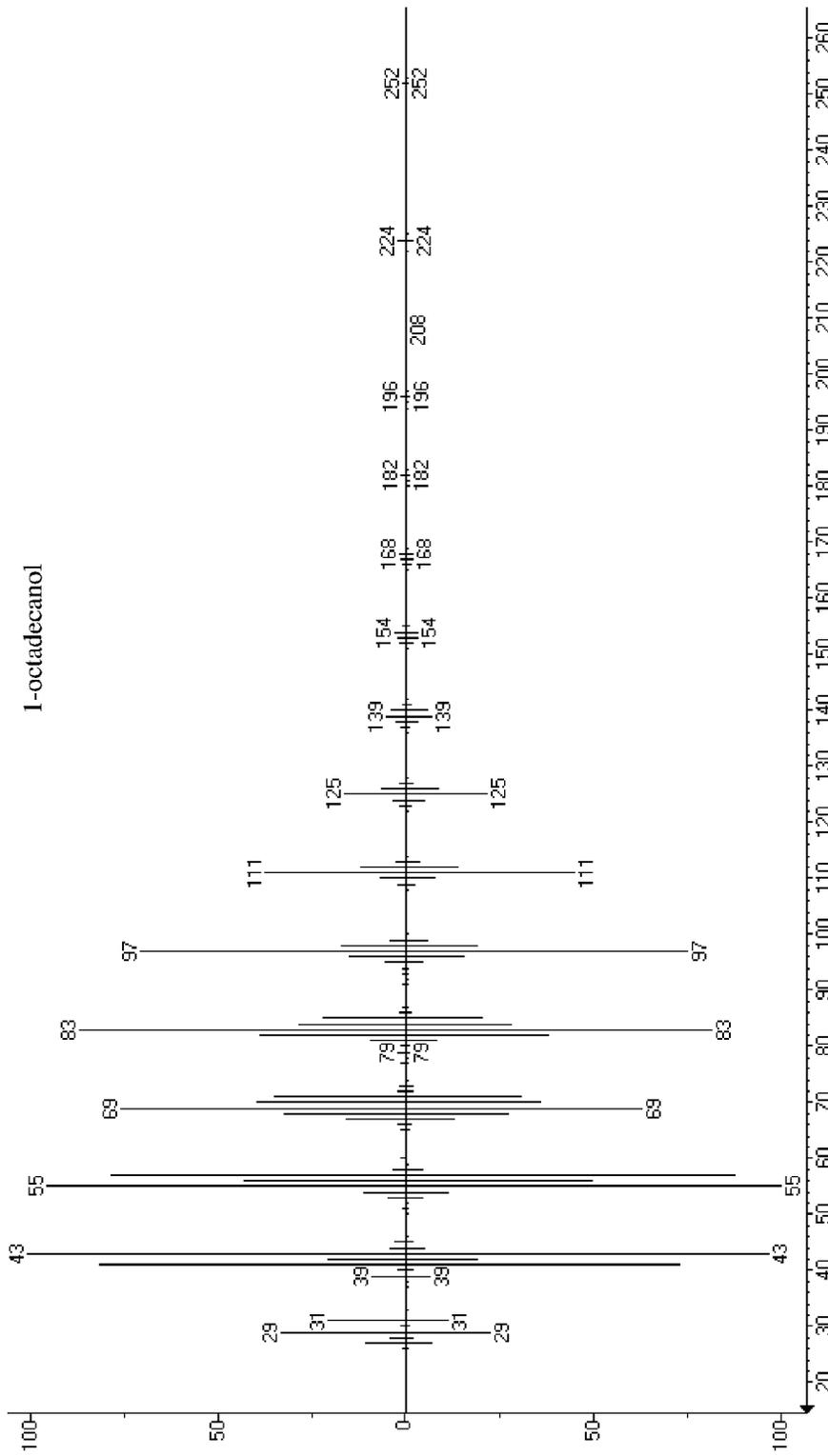


FIG. 9F

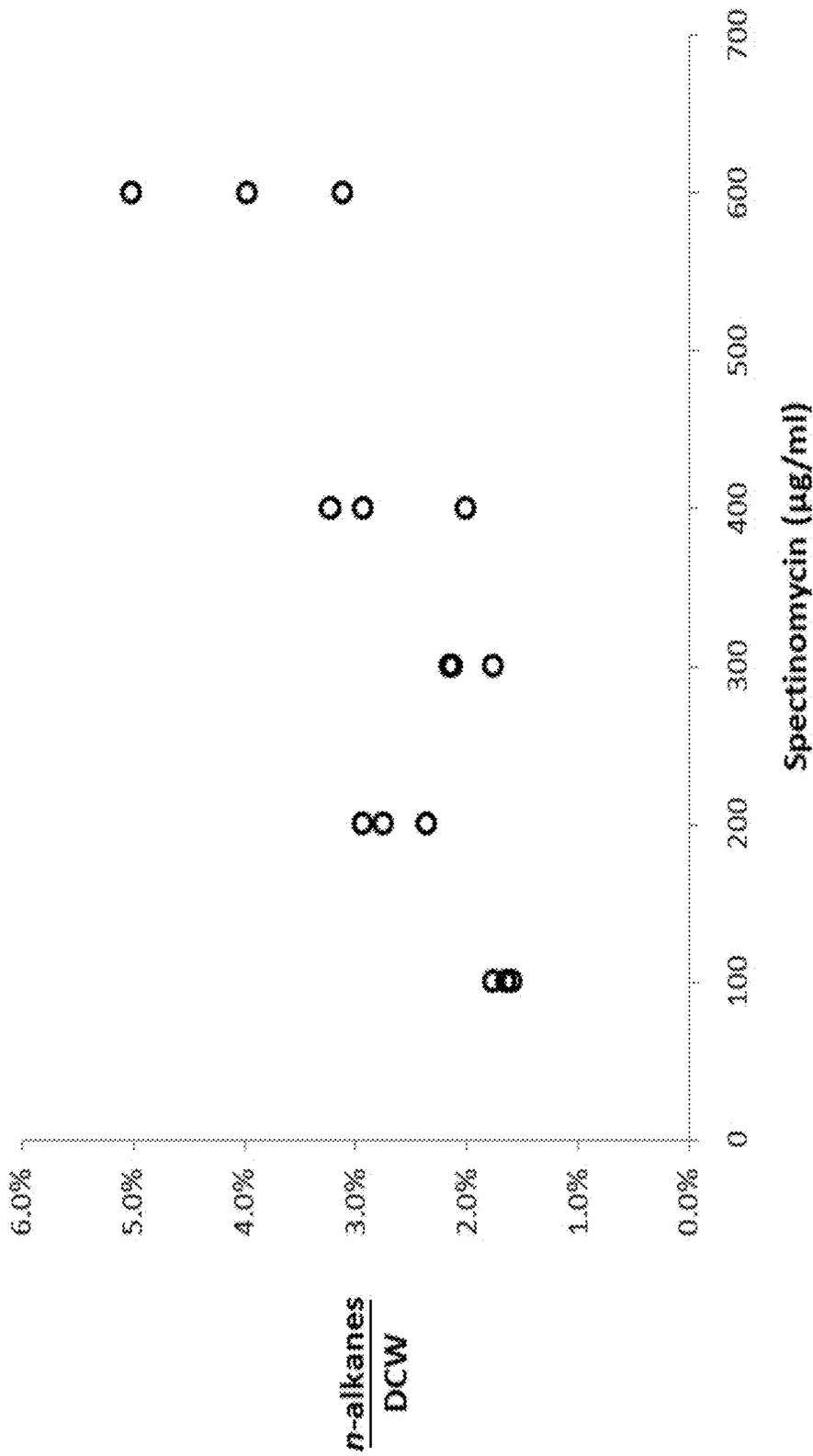


FIG. 10

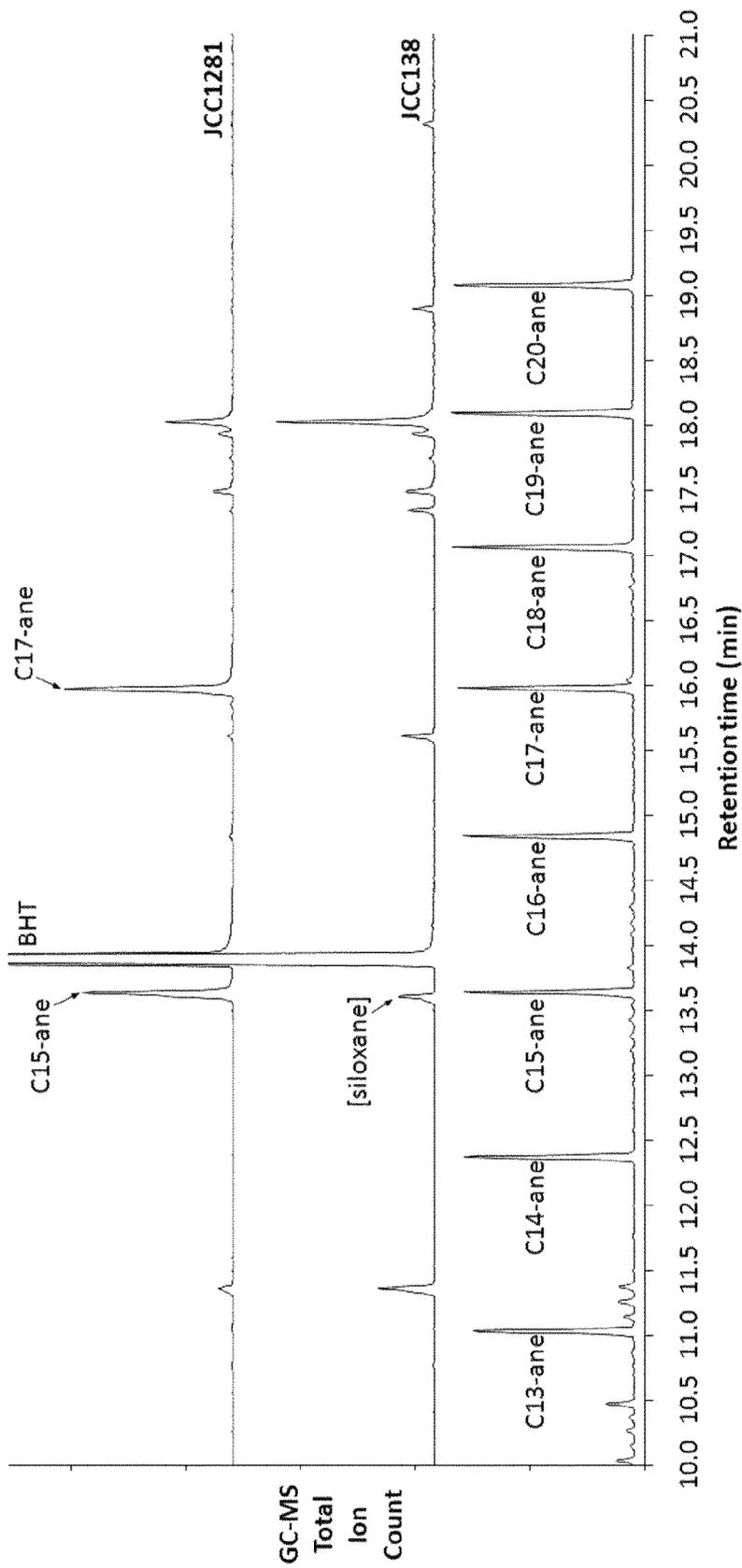


FIG. 11

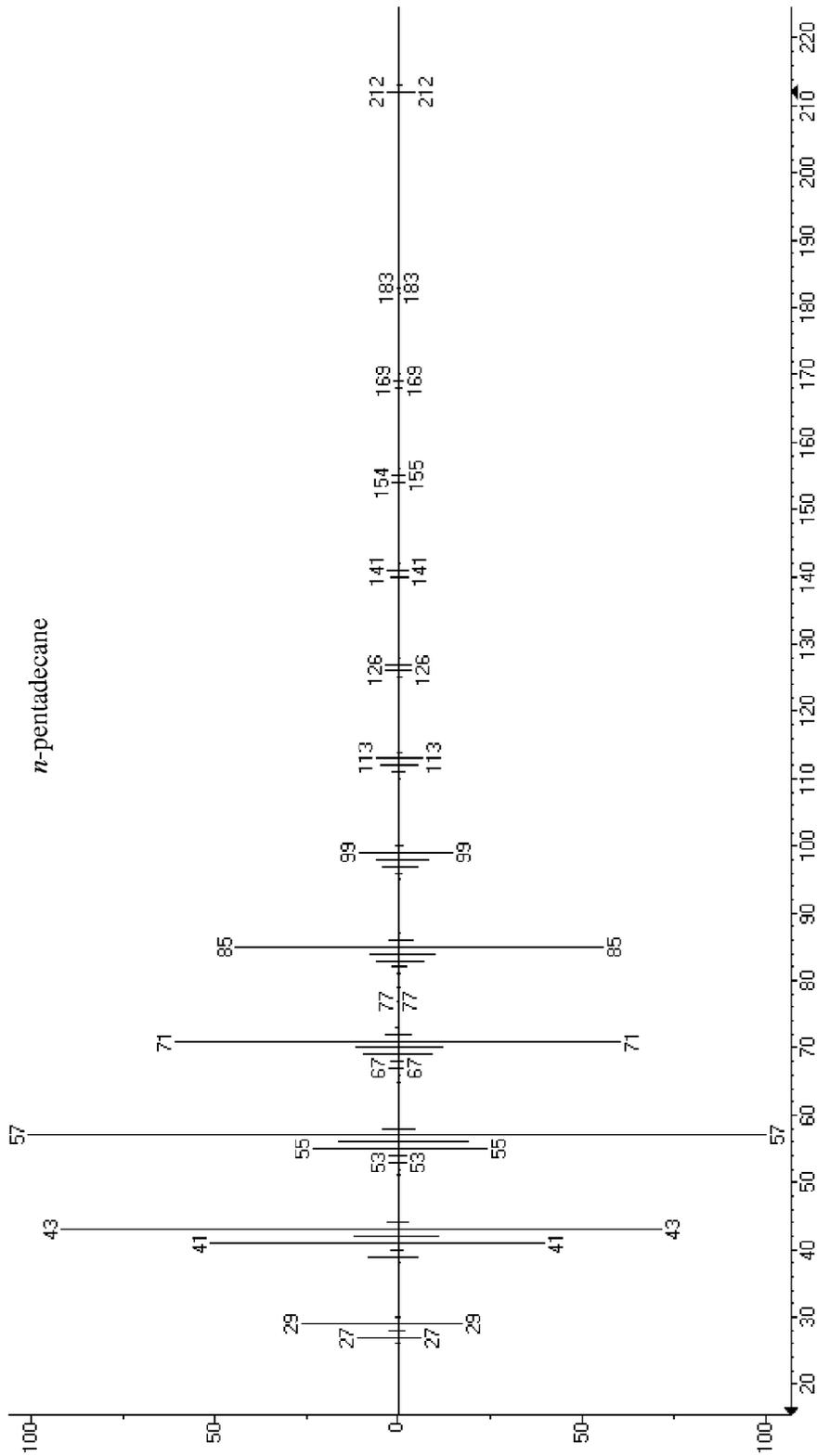


FIG. 12A

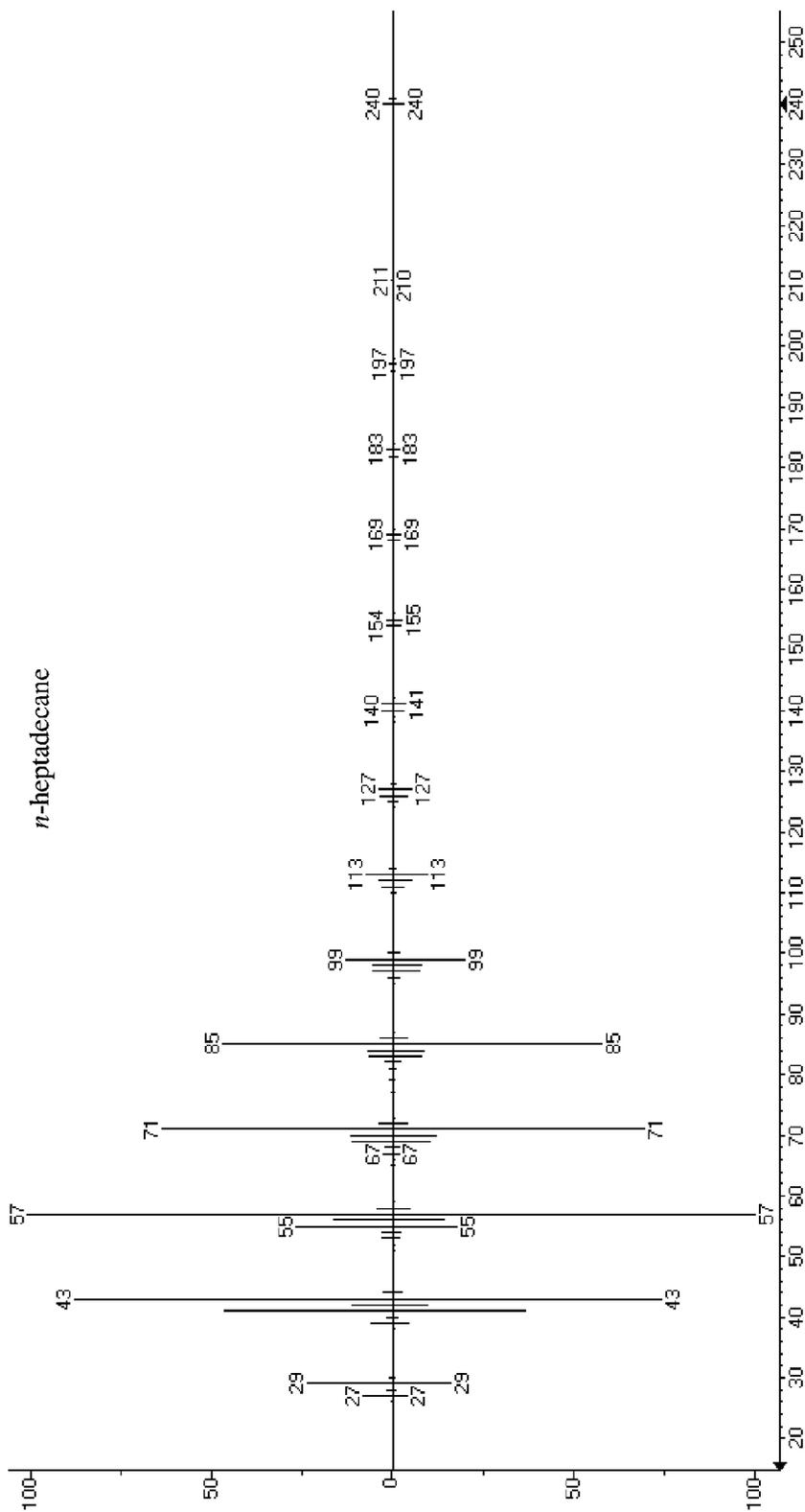


FIG. 12B

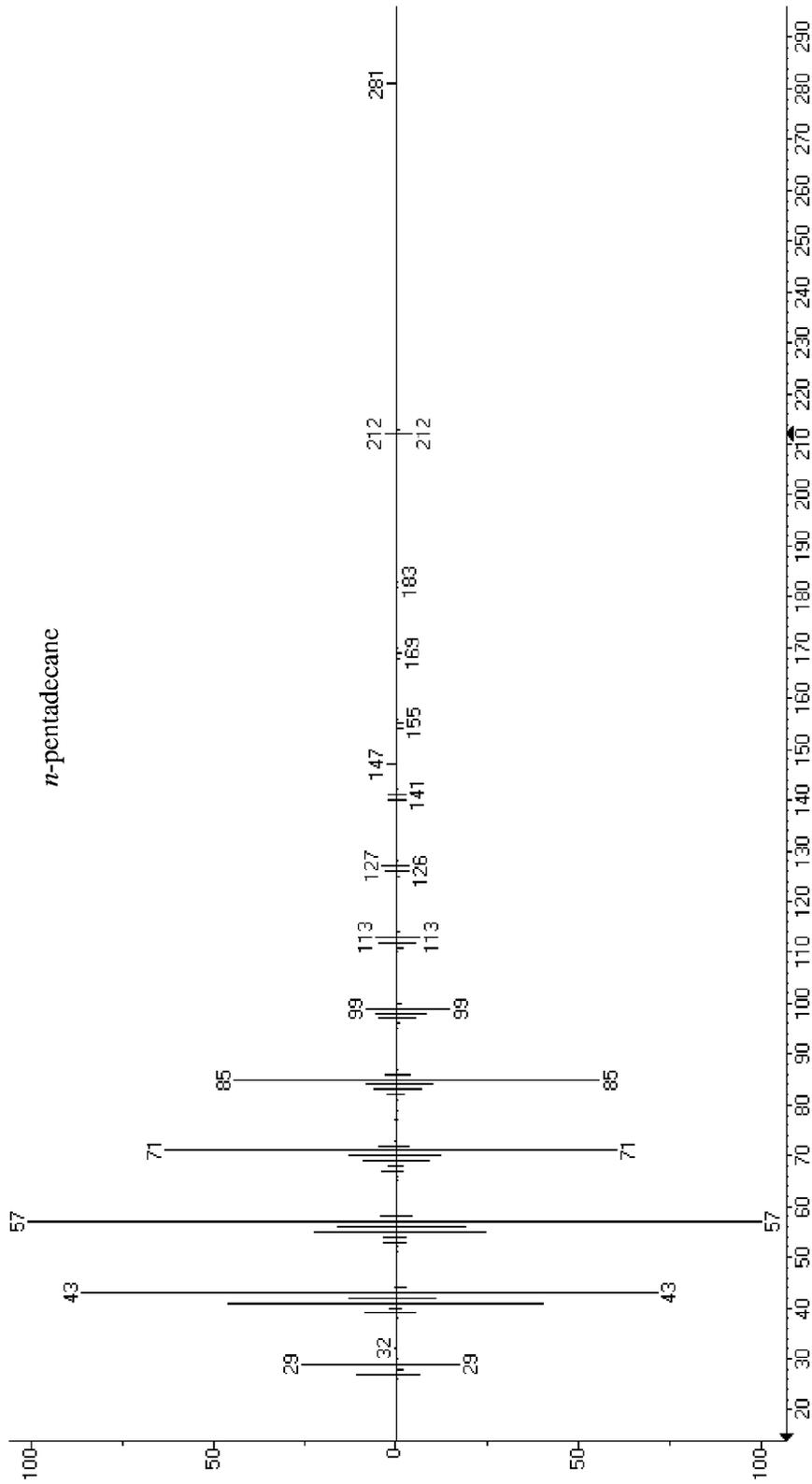


FIG. 13A

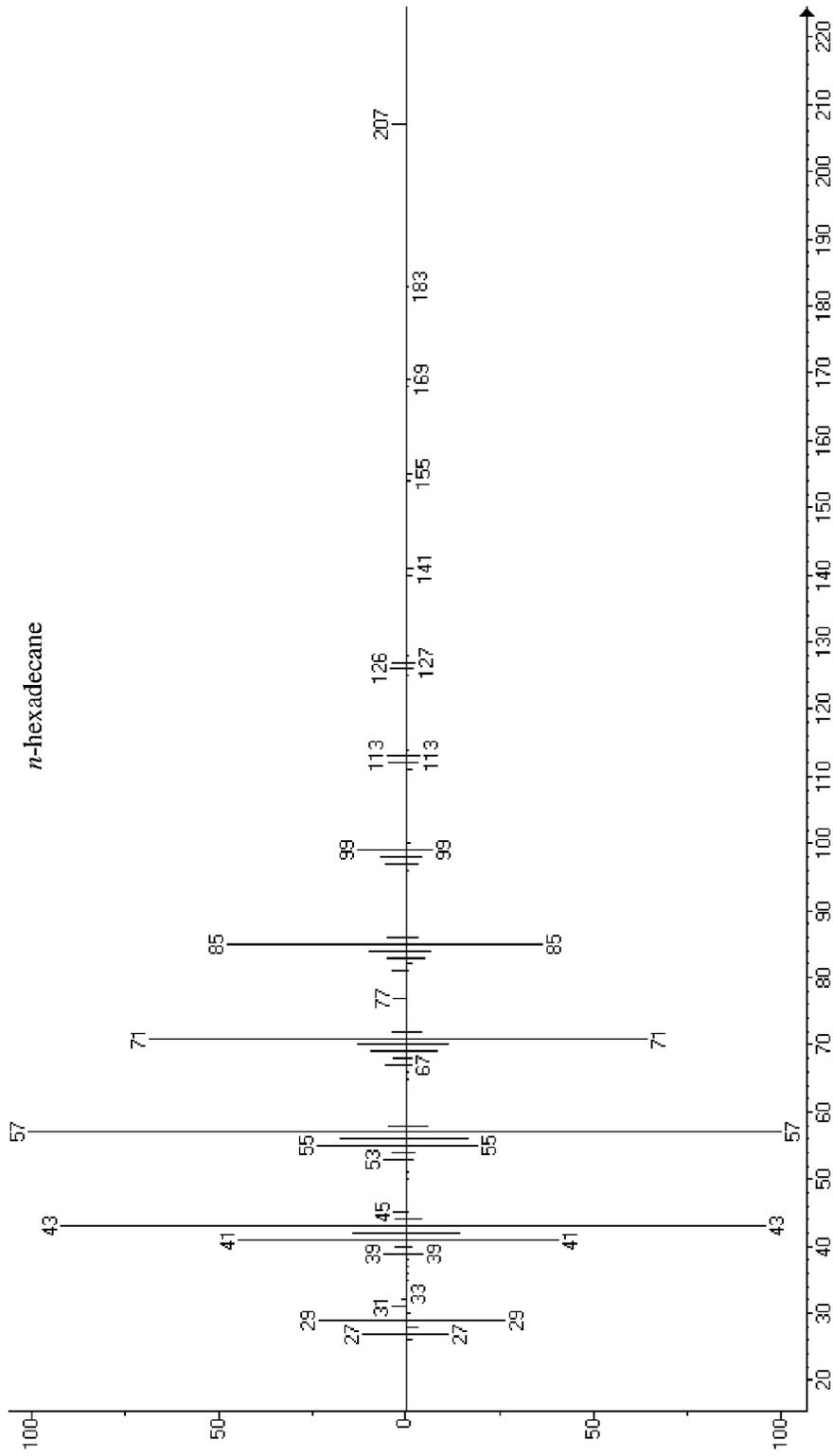


FIG. 13B

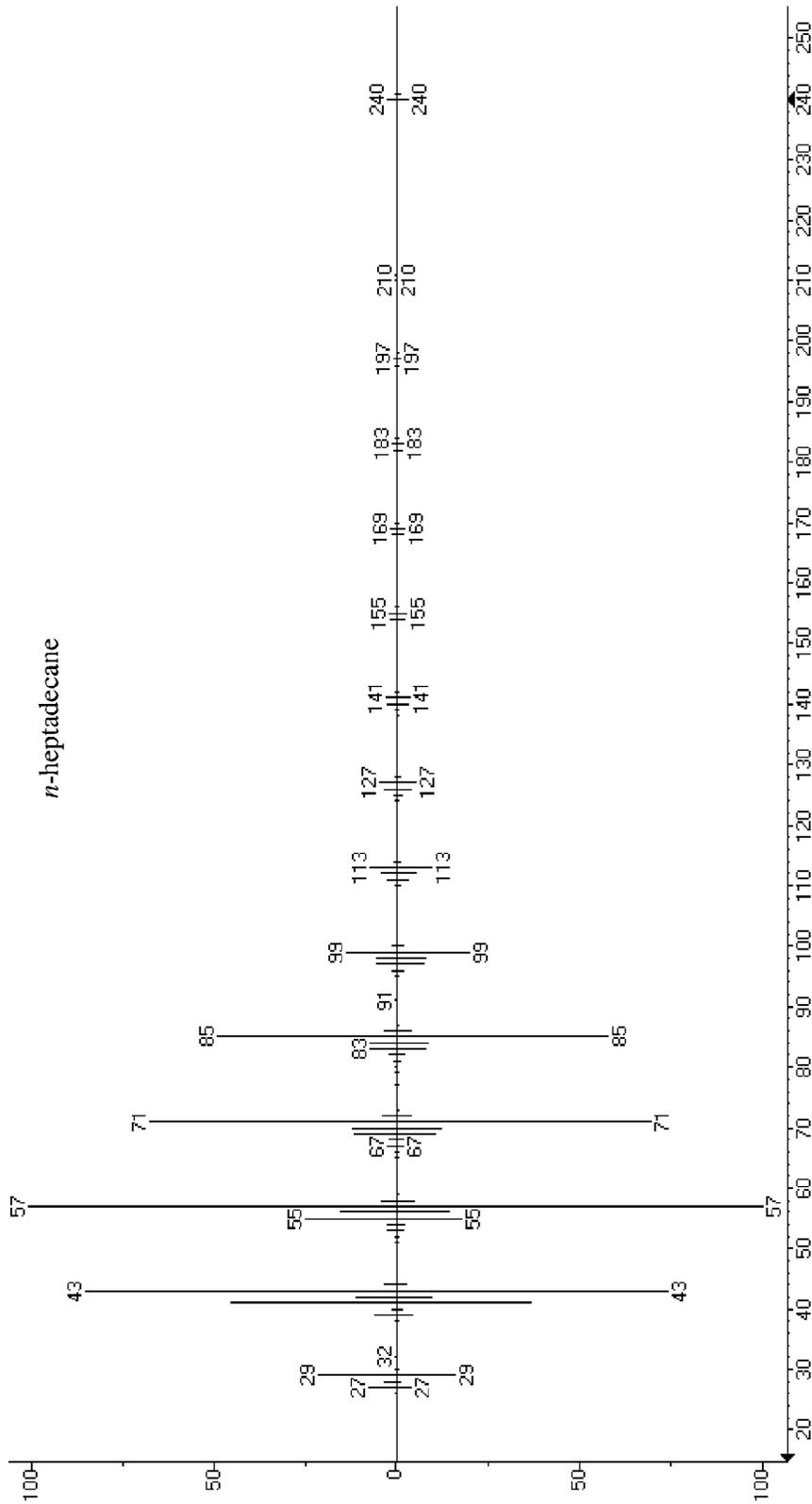


FIG. 13C

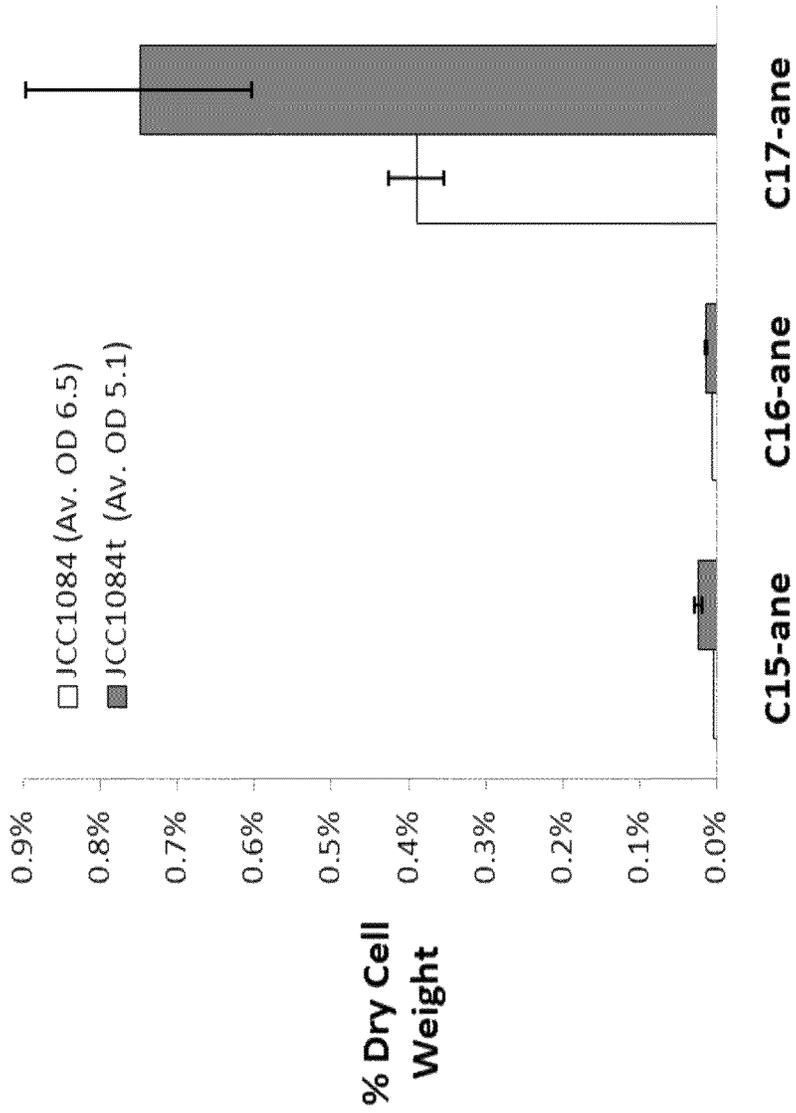


FIG. 14

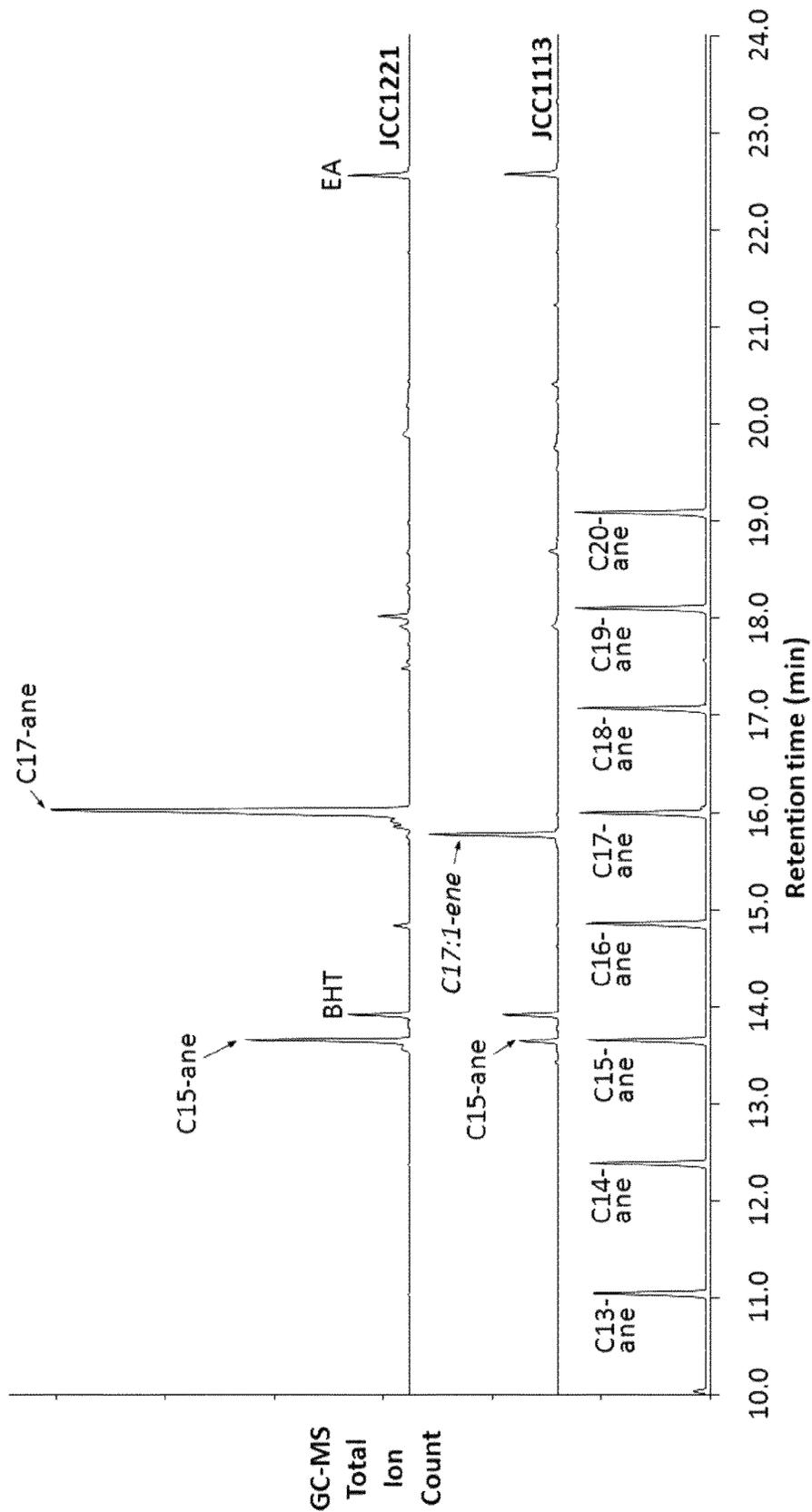


FIG. 15

METHODS AND COMPOSITIONS FOR THE RECOMBINANT BIOSYNTHESIS OF N-ALKANES

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a divisional of U.S. utility application Ser. No. 13/442,840, filed Apr. 9, 2012, which is a continuation of U.S. utility application Ser. No. 13/243,136, filed Sep. 23, 2011, now U.S. Pat. No. 8,183,027, which is a continuation of U.S. utility application Ser. No. 13/098,700, filed May 2, 2011, now U.S. Pat. No. 8,043,840, which is a continuation of U.S. utility application Ser. No. 12/833,821, filed Jul. 9, 2010, now U.S. Pat. No. 7,955,820, which is a continuation-in-part of U.S. utility application Ser. No. 12/759,657, filed Apr. 13, 2010, now U.S. Pat. No. 7,794,969, which claims priority to earlier filed U.S. Provisional Patent Application No. 61/224,463, filed Jul. 9, 2009 and U.S. Provisional Patent Application No. 61/228,937, filed Jul. 27, 2009; the entire disclosures of each of which are incorporated herein by reference, for all purposes.

REFERENCE TO SEQUENCE LISTING

This application is filed with a computer-readable Sequence Listing which has been submitted via EFS-Web and is hereby incorporated by reference in its entirety. Said ASCII copy, created on Apr. 9, 2012, is named "20877_US_Sequence_Listing.txt", lists 128 sequences, and is 332 kb in size.

FIELD OF THE INVENTION

The present disclosure relates to methods for conferring alkane-producing properties to a heterotrophic or photoautotrophic host, such that the modified host can be used in the commercial production of bioalkanes.

BACKGROUND OF THE INVENTION

Many existing photoautotrophic organisms (i.e., plants, algae, and photosynthetic bacteria) are poorly suited for industrial bioprocessing and have therefore not demonstrated commercial viability. Such organisms typically have slow doubling times (3-72 hrs) compared to industrialized heterotrophic organisms such as *Escherichia coli* (20 minutes), reflective of low total productivities. While a desire for the efficient biosynthetic production of fuels has led to the development of photosynthetic microorganisms which produce alkyl esters of fatty acids, a need still exists for methods of producing hydrocarbons, e.g., alkanes, using photosynthetic organisms.

SUMMARY OF THE INVENTION

The present invention provides, in certain embodiments, isolated polynucleotides comprising or consisting of nucleic acid sequences selected from the group consisting of the coding sequences for AAR and ADM enzymes, nucleic acid sequences that are codon-optimized variants of these sequences, and related nucleic acid sequences and fragments.

An AAR enzyme refers to an enzyme with the amino acid sequence of the SYNPC7942_1594 protein (SEQ ID NO: 6) or a homolog thereof, wherein a SYNPC7942_1594 homolog is a protein whose BLAST alignment (i) covers >90% length of SYNPC7942_1594, (ii) covers >90% of

the length of the matching protein, and (iii) has >50% identity with SYNPC7942_1594 (when optimally aligned using the parameters provided herein), and retains the functional activity of SYNPC7942_1594, i.e., the conversion of an acyl-ACP (ACP=acyl carrier protein) to an alkanal. An ADM enzyme refers to an enzyme with the amino acid sequence of the SYNPC7942_1593 protein (SEQ ID NO: 8) or a homolog thereof, wherein a SYNPC7942_1593 homolog is defined as a protein whose amino acid sequence alignment (i) covers >90% length of SYNPC7942_1593, (ii) covers >90% of the length of the matching protein, and (iii) has >50% identity with SYNPC7942_1593 (when aligned using the preferred parameters provided herein), and retains the functional activity of SYNPC7942_1593, i.e., the conversion of an n-alkanal to an (n-1)-alkane. Exemplary AAR and ADM enzymes are listed in Table 1 and Table 2, respectively. Genes encoding AAR or ADM enzymes are referred to herein as AAR genes (aar) or ADM genes (adm), respectively.

Preferred parameters for BLASTp are: Expectation value: 10 (default); Filter: none; Cost to open a gap: 11 (default); Cost to extend a gap: 1 (default); Maximum alignments: 100 (default); Word size: 11 (default); No. of descriptions: 100 (default); Penalty Matrix: BLOWSUM62.

While Applicants refer herein to an alkanal decarboxylative monooxygenase enzyme, Applicants do so without intending to be bound to any particular reaction mechanism unless expressly set forth. For example, whether the enzyme encoded by SYNPC7942_1593 or any other ADM gene carries out a decarboxylase or a decarboxylase reaction does not affect the utility of Applicants' invention, unless expressly set forth herein to the contrary.

The present invention further provides isolated polypeptides comprising or consisting of polypeptide sequences selected from the group consisting of the sequences listed in Table 1 and Table 2, and related polypeptide sequences, fragments and fusions. Antibodies that specifically bind to the isolated polypeptides of the present invention are also contemplated.

The present invention also provides methods for expressing a heterologous nucleic acid sequence encoding AAR and ADM in a host cell lacking catalytic activity for AAR and ADM (thereby conferring n-alkane producing capability in the host cell), or for expressing a nucleic acid encoding AAR and ADM in a host cell which comprises native AAR and/or ADM activity (thereby enhancing n-alkane producing capability in the host cell).

In addition, the present invention provides methods for producing carbon-based products of interest using the AAR and ADM genes, proteins and host cells described herein. For example, in one embodiment the invention provides a method for producing hydrocarbons, comprising: (i) culturing an engineered cyanobacterium in a culture medium, wherein said engineered cyanobacterium comprises a recombinant AAR enzyme and a recombinant ADM enzyme; and (ii) exposing said engineered cyanobacterium to light and carbon dioxide, wherein said exposure results in the conversion of said carbon dioxide by said engineered cyanobacterium into n-alkanes, wherein at least one of said n-alkanes is selected from the group consisting of n-tridecane, n-tetradecane, n-pentadecane, n-hexadecane, and n-heptadecane, and wherein the amount of said n-alkanes produced is between 0.1% and 5% dry cell weight and at least two times the amount produced by an otherwise identical cyanobacterium, cultured under identical conditions, but lacking said recombinant AAR and ADM enzymes.

In a related embodiment, the amount of n-alkanes produced by the engineered cyanobacterium is at least 0.1%,

3

0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, or 1% DCW, and at least two times the amount produced by an otherwise identical cyanobacterium, cultured under identical conditions, but lacking said recombinant AAR and ADM enzymes.

In a related embodiment, at least one of said recombinant enzymes is heterologous with respect to said engineered cyanobacterium. In another embodiment, said cyanobacterium does not synthesize alkanes in the absence of the expression of one or both of the recombinant enzymes. In another embodiment, at least one of said recombinant AAR or ADM enzymes is not heterologous to said engineered cyanobacterium.

In another related embodiment of the method, said engineered cyanobacterium further produces at least one n-alkene or n-alkanol. In yet another embodiment, the engineered cyanobacterium produces at least one n-alkene or n-alkanol selected from the group consisting of n-pentadecene, n-heptadecene, and 1-octadecanol. In a related embodiment, said n-alkanes comprise predominantly n-heptadecane, n-pentadecane or a combination thereof. In a related embodiment, more n-heptadecane and/or n-pentadecane are produced than all other n-alkane products combined. In yet another related embodiment, more n-heptadecane and/or n-pentadecane are produced by the engineered cyanobacterium than any other n-alkane or n-alkene produced by the engineered cyanobacterium. In yet another related embodiment, at least one n-pentadecene produced by said engineered cyanobacterium is selected from the group consisting of cis-3-heptadecene and cis-4-pentadecene. In yet another related embodiment, at least one n-heptadecene produced by said engineered cyanobacterium is selected from the group consisting of cis-4-pentadecene, cis-6-heptadecene, cis-8-heptadecene, cis-9-heptadecene, and cis, cis-heptadec-di-ene.

In yet another related embodiment, the invention further provides a step of isolating at least one n-alkane, n-alkene or n-alkanol from said engineered cyanobacterium or said culture medium. In yet another related embodiment, the engineered cyanobacterium is cultured in a liquid medium. In yet another related embodiment, the engineered cyanobacterium is cultured in a photobioreactor.

In another related embodiment, the AAR and/or ADM enzymes are encoded by a plasmid. In yet another related embodiment, the AAR and/or ADM enzymes are encoded by recombinant genes incorporated into the genome of the engineered cyanobacterium. In yet another related embodiment, the AAR and/or ADM enzymes are encoded by genes which are present in multiple copies in said engineered cyanobacterium. In yet another related embodiment, the recombinant AAR and/or ADM enzymes are encoded by genes which are part of an operon, wherein the expression of said genes is controlled by a single promoter. In yet another related embodiment, the recombinant AAR and/or ADM enzymes are encoded by genes which are expressed independently under the control of separate promoters. In yet another related embodiment, expression of the recombinant AAR and/or ADM enzymes in an engineered cyanobacterium is controlled by a promoter selected from the group consisting of a cI promoter, a cpcB promoter, a lacI-trc promoter, an EM7 promoter, an aphII promoter, a nirA promoter, and a nir07 promoter (referred to herein as "P(nir07)"). In yet another related embodiment, the enzymes are encoded by genes which are part of an operon, wherein the expression of said genes is controlled by one or more inducible promoters. In yet another related embodiment, at least one promoter is a urea-repressible, nitrate-inducible promoter. In yet another related embodiment, the urea-repressible, nitrate-inducible promoter

4

is a nirA-type promoter. In yet another related embodiment, the nirA-type promoter is P(nir07) (SEQ ID NO: 24).

In yet another related embodiment, the cyanobacterium species that is engineered to express recombinant AAR and/or ADM enzymes produces less than approximately 0.01% DCW n-heptadecane or n-pentadecane in the absence of said recombinant AAR and/or ADM enzymes, 0.01% DCW corresponding approximately to the limit of detection of n-heptadecane and n-pentadecane by the gas chromatographic/flame ionization detection methods described herein. In another related embodiment, the engineered cyanobacterium of the method is a thermophile. In yet another related embodiment, the engineered cyanobacterium of the method is selected from the group consisting of an engineered *Synechococcus* sp. PCC7002 and an engineered *Thermosynechococcus elongatus* BP-1.

In yet another related embodiment, the recombinant AAR and/or ADM enzymes are selected from the group of enzymes listed in Table 1 and Table 2, respectively. In yet another related embodiment, the recombinant AAR enzymes are selected from the group consisting of SYNPC7942_1594, tll1312, PMT9312_0533, and cce_1430. In yet another related embodiment, the recombinant ADM enzymes are selected from the group consisting of SYNPC7942_1593, tll1313, PMT9312_0532, and cce_0778.

In yet another related embodiment, the recombinant AAR and ADM enzymes have the amino acid sequences of SEQ ID NO:10 and SEQ ID NO:12, respectively. In certain embodiments, the recombinant AAR and ADM enzymes are encoded by SEQ ID NOs: 9 and 11, respectively. In yet other embodiments, the recombinant AAR and ADM enzymes are encoded by SEQ ID NOs: 26 and 28, respectively, or SEQ ID NOs: 30 and 31 respectively, and have the amino acid sequences of SEQ ID NOs: 27 and 28, respectively. In certain embodiments, the recombinant AAR and ADM enzymes are encoded by SEQ ID NOs: 1 and 3, respectively, and have the amino acid sequences of SEQ NOs: 2 and 4, respectively. In still other embodiments, the recombinant AAR and ADM enzymes are encoded by SEQ ID NOs: 5 and 7, respectively, and have the amino acid sequences of SEQ ID NOs: 6 and 8, respectively.

In yet another related embodiment, the method comprising culturing the engineered cyanobacterium in the presence of an antibiotic, wherein said antibiotic selects for the presence of a recombinant gene encoding an AAR and/or ADM enzyme. In certain embodiments, the antibiotic is spectinomycin or kanamycin. In related embodiments, the amount of spectinomycin in the culture media is between 100 and 700 µg/ml, e.g., 100, 200, 300, 400, 500, 600, or 700 µg/ml of spectinomycin can be added to the culture media. In certain embodiments, the amount of spectinomycin added is about 600 µg/ml, and the amount of n-alkanes produced by the engineered cyanobacterium is at least about 3%, 4% or 5% DCW.

In another embodiment, the method for producing hydrocarbons comprises culturing a cyanobacterium expressing recombinant AAR and/or ADM enzymes in the presence of an exogenous substrate for one or both enzymes. In a related embodiment, the substrate is selected from the group consisting of an acyl-ACP, an acyl-CoA, and a fatty aldehyde. In another related embodiment, exogenous fatty alcohols or fatty esters or other indirect substrates can be added and converted to acyl-ACP or acyl-CoA by the cyanobacterium.

In yet another embodiment, the invention provides a composition comprising an n-alkane produced by any of the recombinant biosynthetic methods described herein. In yet another embodiment, the invention provides a composition

comprising an n-alkene or n-alkanol produced by any of the recombinant biosynthetic methods described herein.

In certain embodiments, the invention provides an engineered host cell for producing an n-alkane, wherein said cell comprises one or more recombinant protein activities selected from the group consisting of an acyl-CoA reductase activity, an acyl-ACP reductase activity, an alkanal decarboxylative monooxygenase activity, and an electron donor activity. In related embodiments, the host cell comprises a recombinant acyl-ACP reductase activity, a recombinant alkanal decarboxylative monooxygenase activity, and a recombinant electron donor activity. In other embodiments, the host cell comprises a recombinant acyl-ACP reductase activity and a recombinant alkanal decarboxylative monooxygenase activity. In certain embodiments, the electron donor activity is a ferredoxin. In certain related embodiments, the host cell is capable of photosynthesis. In still other related embodiments, the host cell is a cyanobacterium. In still other embodiments, the host cell is a gram-negative bacterium, a gram-positive bacterium, or a yeast species.

In other embodiments, the invention provides an isolated or recombinant polynucleotide comprising or consisting of a nucleic acid sequence selected from the group consisting of: (a) SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 14, 30 or 31; (b) a nucleic acid sequence that is a degenerate variant of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 14, 30 or 31; (c) a nucleic acid sequence at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.1%, at least 99.2%, at least 99.3%, at least 99.4%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8% or at least 99.9% identical to SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 14, 30 or 31; (d) a nucleic acid sequence that encodes a polypeptide having the amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 27 or 29; (e) a nucleic acid sequence that encodes a polypeptide at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.1%, at least 99.2%, at least 99.3%, at least 99.4%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8% or at least 99.9% identical to SEQ ID NO: 2, 4, 6, 8, 10, 12, 27 or 29; and (f) a nucleic acid sequence that hybridizes under stringent conditions to SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 14, 30 or 31. In related embodiments, the nucleic acid sequence encodes a polypeptide having acyl-ACP reductase activity or alkanal decarboxylative monooxygenase activity.

In yet another embodiment, the invention provides an isolated, soluble polypeptide with alkanal decarboxylative monooxygenase activity wherein, in certain related embodiments, the polypeptide has an amino acid sequence of one of the proteins listed in Table 2. In related embodiments, the polypeptide has the amino acid sequence identical to, or at least 95% identical to, SEQ ID NO: 4, 8, 12 or 29.

In yet another embodiment, the invention provides a method for synthesizing an n-alkane from an acyl-ACP in vitro, comprising: contacting an acyl-ACP with a recombinant acyl-ACP reductase, wherein said acyl-ACP reductase converts said acyl-ACP to an n-alkanal; then contacting said n-alkanal with a recombinant, soluble alkanal decarboxylative monooxygenase in the presence of an electron donor, wherein said alkanal decarboxylative monooxygenase converts said n-alkanal to an (n-1) alkane. In a related embodiment, the invention provides a method for synthesizing an n-alkane from an n-alkanal in vitro, comprising: contacting said n-alkanal with a recombinant, soluble alkanal decarboxylative monooxygenase in the presence of an electron donor, wherein said alkanal decarboxylative monooxygenase converts said n-alkanal to an (n-1)-alkane. In certain related embodiments, the electron donor is a ferredoxin protein.

In another embodiment, the invention provides engineered cyanobacterial cells comprising recombinant AAR and ADM enzymes, wherein said cells comprise between 0.1% and 5%, between 1% and 5%, or between 2% and 5% dry cell weight n-alkanes, wherein said n-alkanes are predominantly n-pentadecane, n-heptadecane, or a combination thereof.

In other embodiments, the invention provides one of the expression and/or transformation vectors disclosed herein. In other related embodiments, the invention provides methods of using one of the expression and/or transformation vectors disclosed herein to transform a microorganism, e.g., a cyanobacterium.

In yet another embodiment of the method for producing hydrocarbons, the AAR and ADM enzymes are at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO: 6 and SEQ ID NO: 8, respectively. In a related embodiment, the engineered cyanobacterium produces n-pentadecane and n-heptadecane, wherein the percentage by mass of n-pentadecane relative to n-pentadecane plus n-heptadecane is at least 20%. In yet another related embodiment, the engineered cyanobacterium produces n-pentadecane and n-heptadecane, wherein the percentage by mass of n-pentadecane relative to n-pentadecane plus n-heptadecane is less than 30%. In yet another related embodiment, the engineered cyanobacterium produces n-pentadecane and n-heptadecane, wherein the percentage by mass of n-pentadecane relative to n-pentadecane plus n-heptadecane is between 20% and 30%.

In yet another embodiment of the method for producing hydrocarbons, the AAR and ADM enzymes are at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO: 10 and SEQ ID NO: 12, respectively. In a related embodiment, the engineered cyanobacterium produces n-pentadecane and n-heptadecane, wherein the percentage by mass of n-pentadecane relative to n-pentadecane plus n-heptadecane is at least 50%. In yet another related embodiment, the percentage by mass of n-pentadecane relative to n-pentadecane plus n-heptadecane is less than 60%. In yet another related embodiment, the percentage by mass of n-pentadecane relative to n-pentadecane plus n-heptadecane is between 50% and 60%.

In yet another embodiment of the method for producing hydrocarbons, the engineered cyanobacterium comprises at least two distinct recombinant ADM enzymes and at least two distinct recombinant AAR enzymes. In a related embodiment, said engineered cyanobacterium comprises at least one operon encoding AAR and ADM enzymes which are at least 95% identical to SEQ ID NO: 27 and SEQ ID NO: 29, respectively. In yet another related embodiment, said engineered cyanobacterium comprises at least one operon encoding AAR and ADM enzymes which are at least 95% identical to SEQ ID NO: 10 and SEQ ID NO: 12, respectively. In yet another related embodiment, expression of said AAR and ADM enzymes is controlled by an inducible promoter, e.g., a P(nir07) promoter. In yet another related embodiment, said recombinant ADM and AAR enzymes are chromosomally integrated. In yet another related embodiment, said engineered cyanobacterium produces n-alkanes in the presence of an inducer, and wherein at least 95% of said n-alkanes are n-pentadecane and n-heptadecane, and wherein the percentage by mass of n-pentadecane relative to n-pentadecane plus n-heptadecane is at least 80%.

In yet another embodiment of the method for producing hydrocarbons, the engineered cyanobacterium comprises recombinant AAR and ADM enzymes which are at least 95% identical to SEQ ID NO: 10 and SEQ ID NO: 12, respectively. In a related embodiment, the recombinant AAR and ADM

enzymes are under the control of an inducible promoter, e.g., a P(nir07) promoter. In yet another related embodiment, the engineered cyanobacterium produces at least 0.5% DCW n-alkanes in the presence of an inducer, and wherein said n-alkanes comprise n-pentadecane and n-heptadecane, and wherein the percentage by mass of n-pentadecane relative to n-pentadecane plus n-heptadecane is at least 50%.

In yet another embodiment, the invention provides a method for modulating the relative amounts of n-pentadecane and n-heptadecane in an engineered cyanobacterium, comprising controlling the expression of one or more recombinant AAR and/or ADM enzymes in said cyanobacterium, wherein said AAR and/or ADM enzymes are at least 95% identical or identical to the AAR and ADM enzymes of SEQ ID NO:s 10, 12, 27 or 29.

In another embodiment, the invention provides an engineered cyanobacterium, wherein said engineered cyanobacterium comprises one or more recombinant genes encoding an AAR enzyme, an ADM enzyme, or both enzymes, wherein at least one of said recombinant genes is under the control of a nitrate-inducible promoter.

In yet another embodiment, the invention provides a recombinant gene, wherein said gene comprises a promoter for controlling expression of said gene, wherein said promoter comprises a contiguous nucleic acid sequence identical to SEQ ID NO: 24.

In yet another embodiment, the invention provides an isolated DNA molecule comprising a promoter, wherein said promoter comprises a contiguous nucleic acid sequence identical to SEQ ID NO: 24.

In yet another embodiment, the invention provides an engineered bacterial strain selected from the group consisting of JCC1469, JCC1281, JCC1683, JCC1685, JCC1076, JCC1170, JCC1221, JCC879 and JCC1084t.

These and other embodiments of the invention are further described in the Figures, Description, Examples and Claims, herein.

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1 depicts, in panel FIG. 1A, an enzymatic pathway for the production of n-alkanes based on the sequential activity of (1) an AAR enzyme (e.g., tll1312); and (2) an ADM enzyme (e.g., tll1313); FIG. 1B, Biosynthesis of n-alkanal via acyl-CoA. Acyl-CoAs are typically intermediates of fatty acid degradation; FIG. 1C, Biosynthesis of n-alkanal via acyl-ACP. Acyl-ACP's are typically intermediates of fatty acid biosynthesis. Note the three different types of ACP reductase: (i) β -ketoacyl-ACP reductase, (ii) enoyl-ACP reductase, and (iii) acyl-ACP reductase. Acyl-ACP reductase, a new enzyme, generates the substrate for alkanal decarboxylative monooxygenase. CoA, coenzyme A; ACP, acyl carrier protein; FIG. 1D, an alternative acyl-CoA-mediated alkane biosynthetic pathway. See additional discussion in Example 1, herein.

FIG. 2 represents 0-to-2700000-count total ion chromatograms of JCC9a and JCC1076 BHT (butylated hydroxytoluene)-acetone cell pellet extracts, as well as n-alkane and n-1-alkanol authentic standards. Peaks assigned by Method 1 are identified in regular font, those by Method 2 in italic font.

FIG. 3 depicts MS fragmentation spectra of JCC1076 peaks assigned by Method 1 (top mass spectrum of each panel), plotted against their respective NIST library hits (bottom mass spectrum of each panel). FIG. 3A, n-pentadecane; FIG. 3B, 1-tetradecanol; FIG. 3C, n-heptadecane; FIG. 3D, 1-hexadecanol.

FIG. 4A represents 0-to-7500000-count total ion chromatograms for the BHT-acetone extracts of JCC1113 and JCC1114 cell pellets, as well as C₁₃-C₂₀ n-alkane and C₁₄, C₁₆, and C₁₈ n-1-alkanol authentic standards; FIG. 4B, represents 0-to-720000-count total ion chromatograms for BHT-acetone extracts of JCC1113 and JCC1114 cell pellets, as well as the n-alkane and n-alkanol authentic standards mentioned in FIG. 4A.

FIG. 5 depicts MS fragmentation spectra of JCC1113 peaks assigned by Method 1 (top mass spectrum of each panel), plotted against their respective NIST library hits (bottom mass spectrum of each panel). FIG. 5A, n-tridecane; FIG. 5B, n-tetradecane; FIG. 5C, n-pentadecane; FIG. 5D, n-hexadecane; FIG. 5E, n-heptadecane; FIG. 5F, 1-hexadecanol.

FIG. 6 represents 0-to-6100000-count total ion chromatograms of JCC1170 and JCC1169 BHT-acetone cell pellet extracts versus those of the control strains JCC1113 and JCC1114. No hydrocarbon products were observed in JCC1169. The unidentified peak in JCC1170 is likely cis-11-octadecenal.

FIG. 7 depicts MS fragmentation spectra of JCC1170 peaks assigned by Method 1 (top mass spectrum of each panel), plotted against their respective NIST library hits (bottom mass spectrum of each panel). FIG. 7A, 1-tetradecanol; FIG. 7B, 1-hexadecanol.

FIG. 8A represents 0-to-7500000-count total ion chromatograms for BHT-acetone extracts of JCC1221, JCC1220, JCC1160b, JCC1160a, JCC1160 and JCC879 cell pellets, as well as C₁₃-C₂₀ n-alkane and C₁₄, C₁₆, and C₁₈ n-alkanol authentic standards. The doublet around 18.0 minutes corresponds to nonadec-di-ene and 1-nonadecene, respectively (data not shown), n-alkenes that are naturally produced by JCC138; FIG. 8B represents 0-to-2250000-count total ion chromatograms for BHT-acetone extracts of JCC1221 and JCC879 cell pellets, as well as the n-alkane and n-alkanol authentic standards mentioned in FIG. 8A.

FIG. 9 depicts MS fragmentation spectra of JCC1221 peaks assigned by Method 1 (top mass spectrum of each panel), plotted against their respective NIST library hits (bottom mass spectrum of each panel). FIG. 9A, n-tridecane; FIG. 9B, n-tetradecane; FIG. 9C, n-pentadecane; FIG. 9D, n-hexadecane; FIG. 9E, n-heptadecane; FIG. 9F, 1-octadecanol.

FIG. 10 depicts intracellular n-alkane production as a function of spectinomycin concentration in JCC1221.

FIG. 11 represents 0-to-1080000-count total ion chromatograms of the JCC1281 BHT-acetone cell pellet extractant versus that of the control strain JCC138, as well as of authentic standard n-alkanes.

FIG. 12 depicts MS fragmentation spectra of JCC1281 peaks assigned by Method 1 (top mass spectrum of each panel), plotted against their respective NIST library hits (bottom mass spectrum of each panel). FIG. 12A, n-pentadecane; FIG. 12B, n-heptadecane.

FIG. 13 depicts MS fragmentation spectra of JCC3 peaks assigned by Method 1 (top mass spectrum of each panel), plotted against their respective NIST library hits (bottom mass spectrum of each panel). FIG. 13A, n-pentadecane; FIG. 13B, n-hexadecane; FIG. 13C, n-heptadecane.

FIG. 14 depicts enhanced intracellular production of n-alkanes in JCC1084t compared to the control strain JCC1084. Error bars represent standard deviation of three independent observations.

FIG. 15 represents 0-to-3150000-count total ion chromatograms of JCC1113 and JCC1221 BHT-acetone cell pellet extracts, as well as authentic n-alkane standards.

DETAILED DESCRIPTION OF THE INVENTION

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have

the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include the plural and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, biochemistry, enzymology, molecular and cellular biology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well known and commonly used in the art.

The methods and techniques of the present invention are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. See, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989); Ausubel et al., *Current Protocols in Molecular Biology*, Greene Publishing Associates (1992, and Supplements to 2002); Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1990); Taylor and Drickamer, *Introduction to Glycobiology*, Oxford Univ. Press (2003); *Worthington Enzyme Manual*, Worthington Biochemical Corp., Freehold, N.J.; *Handbook of Biochemistry: Section A Proteins*, Vol I, CRC Press (1976); *Handbook of Biochemistry: Section A Proteins*, Vol II, CRC Press (1976); *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press (1999).

All publications, patents and other references mentioned herein are hereby incorporated by reference in their entireties.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

The term “polynucleotide” or “nucleic acid molecule” refers to a polymeric form of nucleotides of at least 10 bases in length. The term includes DNA molecules (e.g., cDNA or genomic or synthetic DNA) and RNA molecules (e.g., mRNA or synthetic RNA), as well as analogs of DNA or RNA containing non-natural nucleotide analogs, non-native internucleoside bonds, or both. The nucleic acid can be in any topological conformation. For instance, the nucleic acid can be single-stranded, double-stranded, triple-stranded, quadruplexed, partially double-stranded, branched, hairpinned, circular, or in a padlocked conformation.

Unless otherwise indicated, and as an example for all sequences described herein under the general format “SEQ ID NO:”, “nucleic acid comprising SEQ ID NO:1” refers to a nucleic acid, at least a portion of which has either (i) the sequence of SEQ ID NO:1, or (ii) a sequence complementary to SEQ ID NO:1. The choice between the two is dictated by the context. For instance, if the nucleic acid is used as a probe, the choice between the two is dictated by the requirement that the probe be complementary to the desired target.

An “isolated” RNA, DNA or a mixed polymer is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, e.g., ribosomes, polymerases and genomic sequences with which it is naturally associated.

As used herein, an “isolated” organic molecule (e.g., an alkane, alkene, or alkanal) is one which is substantially separated from the cellular components (membrane lipids, chromosomes, proteins) of the host cell from which it originated, or from the medium in which the host cell was cultured. The term does not require that the biomolecule has been separated from all other chemicals, although certain isolated biomolecules may be purified to near homogeneity.

The term “recombinant” refers to a biomolecule, e.g., a gene or protein, that (1) has been removed from its naturally

occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the gene is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, or (4) does not occur in nature. The term “recombinant” can be used in reference to cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems, as well as proteins and/or mRNAs encoded by such nucleic acids.

As used herein, an endogenous nucleic acid sequence in the genome of an organism (or the encoded protein product of that sequence) is deemed “recombinant” herein if a heterologous sequence is placed adjacent to the endogenous nucleic acid sequence, such that the expression of this endogenous nucleic acid sequence is altered. In this context, a heterologous sequence is a sequence that is not naturally adjacent to the endogenous nucleic acid sequence, whether or not the heterologous sequence is itself endogenous (originating from the same host cell or progeny thereof) or exogenous (originating from a different host cell or progeny thereof). By way of example, a promoter sequence can be substituted (e.g., by homologous recombination) for the native promoter of a gene in the genome of a host cell, such that this gene has an altered expression pattern. This gene would now become “recombinant” because it is separated from at least some of the sequences that naturally flank it.

A nucleic acid is also considered “recombinant” if it contains any modifications that do not naturally occur to the corresponding nucleic acid in a genome. For instance, an endogenous coding sequence is considered “recombinant” if it contains an insertion, deletion or a point mutation introduced artificially, e.g., by human intervention. A “recombinant nucleic acid” also includes a nucleic acid integrated into a host cell chromosome at a heterologous site and a nucleic acid construct present as an episome.

As used herein, the phrase “degenerate variant” of a reference nucleic acid sequence encompasses nucleic acid sequences that can be translated, according to the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence. The term “degenerate oligonucleotide” or “degenerate primer” is used to signify an oligonucleotide capable of hybridizing with target nucleic acid sequences that are not necessarily identical in sequence but that are homologous to one another within one or more particular segments.

The term “percent sequence identity” or “identical” in the context of nucleic acid sequences refers to the residues in the two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wis. FASTA provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences. Pearson, *Methods Enzymol.* 183:63-98 (1990) (hereby incorporated by reference in its entirety). For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as

provided in GCG Version 6.1, herein incorporated by reference. Alternatively, sequences can be compared using the computer program, BLAST (Altschul et al., *J. Mol. Biol.* 215:403-410 (1990); Gish and States, *Nature Genet.* 3:266-272 (1993); Madden et al., *Meth. Enzymol.* 266:131-141 (1996); Altschul et al., *Nucleic Acids Res.* 25:3389-3402 (1997); Zhang and Madden, *Genome Res.* 7:649-656 (1997)), especially blastp or tblastn (Altschul et al., *Nucleic Acids Res.* 25:3389-3402 (1997)).

The term "substantial homology" or "substantial similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 76%, 80%, 85%, preferably at least about 90%, and more preferably at least about 95%, 96%, 97%, 98% or 99% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial homology or similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under stringent hybridization conditions. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization.

In general, "stringent hybridization" is performed at about 25° C. below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5° C. lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. See Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), page 9.51, hereby incorporated by reference. For purposes herein, "stringent conditions" are defined for solution phase hybridization as aqueous hybridization (i.e., free of formamide) in 6×SSC (where 20×SSC contains 3.0 M NaCl and 0.3 M sodium citrate), 1% SDS at 65° C. for 8-12 hours, followed by two washes in 0.2×SSC, 0.1% SDS at 65° C. for 20 minutes. It will be appreciated by the skilled worker that hybridization at 65° C. will occur at different rates depending on a number of factors including the length and percent identity of the sequences which are hybridizing.

The nucleic acids (also referred to as polynucleotides) of this present invention may include both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. They may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), pendent moieties (e.g., polypep-

tides), intercalators (e.g., acridine, psoralen, etc.), chelators, alkylators, and modified linkages (e.g., alpha anomeric nucleic acids, etc.) Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule. Other modifications can include, for example, analogs in which the ribose ring contains a bridging moiety or other structure such as the modifications found in "locked" nucleic acids.

The term "mutated" when applied to nucleic acid sequences means that nucleotides in a nucleic acid sequence may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. A nucleic acid sequence may be mutated by any method known in the art including but not limited to mutagenesis techniques such as "error-prone PCR" (a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product; see, e.g., Leung et al., *Technique*, 1:11-15 (1989) and Caldwell and Joyce, *PCR Methods Applic.* 2:28-33 (1992)); and "oligonucleotide-directed mutagenesis" (a process which enables the generation of site-specific mutations in any cloned DNA segment of interest; see, e.g., Reidhaar-Olson and Sauer, *Science* 241:53-57 (1988)).

The term "attenuate" as used herein generally refers to a functional deletion, including a mutation, partial or complete deletion, insertion, or other variation made to a gene sequence or a sequence controlling the transcription of a gene sequence, which reduces or inhibits production of the gene product, or renders the gene product non-functional. In some instances a functional deletion is described as a knockout mutation. Attenuation also includes amino acid sequence changes by altering the nucleic acid sequence, placing the gene under the control of a less active promoter, down-regulation, expressing interfering RNA, ribozymes or antisense sequences that target the gene of interest, or through any other technique known in the art. In one example, the sensitivity of a particular enzyme to feedback inhibition or inhibition caused by a composition that is not a product or a reactant (non-pathway specific feedback) is lessened such that the enzyme activity is not impacted by the presence of a compound. In other instances, an enzyme that has been altered to be less active can be referred to as attenuated.

Deletion: The removal of one or more nucleotides from a nucleic acid molecule or one or more amino acids from a protein, the regions on either side being joined together.

Knock-out: A gene whose level of expression or activity has been reduced to zero. In some examples, a gene is knocked-out via deletion of some or all of its coding sequence. In other examples, a gene is knocked-out via introduction of one or more nucleotides into its open reading frame, which results in translation of a non-sense or otherwise non-functional protein product.

The term "vector" as used herein is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid," which generally refers to a circular double stranded DNA loop into which additional DNA segments may be ligated, but also includes linear double-stranded molecules such as those resulting from amplification by the polymerase chain reaction (PCR) or from treatment of a circular

plasmid with a restriction enzyme. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome (discussed in more detail below). Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., vectors having an origin of replication which functions in the host cell). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and are thereby replicated along with the host genome. Moreover, certain preferred vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as “recombinant expression vectors” (or simply “expression vectors”).

“Operatively linked” or “operably linked” expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in trans or at a distance to control the gene of interest.

The term “expression control sequence” as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (e.g., ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence. The term “control sequences” is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

The term “recombinant host cell” (or simply “host cell”), as used herein, is intended to refer to a cell into which a recombinant vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term “host cell” as used herein. A recombinant host cell may be an isolated cell or cell line grown in culture or may be a cell which resides in a living tissue or organism.

The term “peptide” as used herein refers to a short polypeptide, e.g., one that is typically less than about 50 amino acids long and more typically less than about 30 amino acids long. The term as used herein encompasses analogs and mimetics that mimic structural and thus biological function.

The term “polypeptide” encompasses both naturally-occurring and non-naturally-occurring proteins, and fragments, mutants, derivatives and analogs thereof. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different domains each of which has one or more distinct activities.

The term “isolated protein” or “isolated polypeptide” is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) exists in a purity not found in nature, where purity can be adjudged with respect to the presence of other cellular material (e.g., is free of other proteins from the same species) (3) is expressed by a cell from a different species, or (4) does not occur in nature (e.g., it is a fragment of a polypeptide found in nature or it includes amino acid analogs or derivatives not found in nature or linkages other than standard peptide bonds). Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be “isolated” from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well known in the art. As thus defined, “isolated” does not necessarily require that the protein, polypeptide, peptide or oligopeptide so described has been physically removed from its native environment.

The term “polypeptide fragment” as used herein refers to a polypeptide that has a deletion, e.g., an amino-terminal and/or carboxy-terminal deletion compared to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, even more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A “modified derivative” refers to polypeptides or fragments thereof that are substantially homologous in primary structural sequence but which include, e.g., in vivo or in vitro chemical and biochemical modifications or which incorporate amino acids that are not found in the native polypeptide. Such modifications include, for example, acetylation, carboxylation, phosphorylation, glycosylation, ubiquitination, labeling, e.g., with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well known in the art, and include radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H , ligands which bind to labeled antiligands (e.g., antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well known in the art. See, e.g., Ausubel et al., *Current Protocols in Molecular Biology*, Greene Publishing Associates (1992, and Supplements to 2002) (hereby incorporated by reference).

The term “fusion protein” refers to a polypeptide comprising a polypeptide or fragment coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusions that include the entirety of the proteins of the present invention have particular utility. The heterologous polypep-

tide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger polypeptides, such as an IgG Fc region, and even entire proteins, such as the green fluorescent protein ("GFP") chromophore-containing proteins, have particular utility. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

As used herein, the term "antibody" refers to a polypeptide, at least a portion of which is encoded by at least one immunoglobulin gene, or fragment thereof, and that can bind specifically to a desired target molecule. The term includes naturally-occurring forms, as well as fragments and derivatives.

Fragments within the scope of the term "antibody" include those produced by digestion with various proteases, those produced by chemical cleavage and/or chemical dissociation and those produced recombinantly, so long as the fragment remains capable of specific binding to a target molecule. Among such fragments are Fab, Fab', Fv, F(ab')₂, and single chain Fv (scFv) fragments.

Derivatives within the scope of the term include antibodies (or fragments thereof) that have been modified in sequence, but remain capable of specific binding to a target molecule, including: interspecies chimeric and humanized antibodies; antibody fusions; heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies (see, e.g., *Intracellular Antibodies: Research and Disease Applications*, (Marasco, ed., Springer-Verlag New York, Inc., 1998), the disclosure of which is incorporated herein by reference in its entirety).

As used herein, antibodies can be produced by any known technique, including harvest from cell culture of native B lymphocytes, harvest from culture of hybridomas, recombinant expression systems and phage display.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." See, e.g., Jones, *Amino Acid and Peptide Synthesis*, Oxford University Press (1992); Jung, *Combinatorial Peptide and Nonpeptide Libraries: A Handbook*, John Wiley (1997); Bodanszky et al., *Peptide Chemistry—A Practical Textbook*, Springer Verlag (1993); *Synthetic Peptides: A Users Guide*, (Grant, ed., W. H. Freeman and Co., 1992); Evans et al., *J. Med. Chem.* 30:1229 (1987); Fauchere, *J. Adv. Drug Res.* 15:29 (1986); Veber and Freidinger, *Trends Neurosci.*, 8:392-396 (1985); and references cited in each of the above, which are incorporated herein by reference. Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides of the present invention may be used to produce an equivalent effect and are therefore envisioned to be part of the present invention.

A "polypeptide mutant" or "mutein" refers to a polypeptide whose sequence contains an insertion, duplication, deletion, rearrangement or substitution of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the natu-

rally-occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. A mutein may have the same but preferably has a different biological activity compared to the naturally-occurring protein.

A mutein has at least 85% overall sequence homology to its wild-type counterpart. Even more preferred are muteins having at least 90% overall sequence homology to the wild-type protein.

In an even more preferred embodiment, a mutein exhibits at least 95% sequence identity, even more preferably 98%, even more preferably 99% and even more preferably 99.9% overall sequence identity.

Sequence homology may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

Amino acid substitutions can include those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs.

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. See *Immunology—A Synthesis* (Golub and Gren eds., Sinauer Associates, Sunderland, Mass., 2nd ed. 1991), which is incorporated herein by reference. Stereoisomers (e.g., D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α -, α -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ -carboxyglutamate, ϵ -N,N,N-trimethyllysine, ϵ -N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, N-methylarginine, and other similar amino acids and imino acids (e.g., 4-hydroxyproline). In the polypeptide notation used herein, the left-hand end corresponds to the amino terminal end and the right-hand end corresponds to the carboxy-terminal end, in accordance with standard usage and convention.

A protein has "homology" or is "homologous" to a second protein if the nucleic acid sequence that encodes the protein has a similar sequence to the nucleic acid sequence that encodes the second protein. Alternatively, a protein has homology to a second protein if the two proteins have "similar" amino acid sequences. (Thus, the term "homologous proteins" is defined to mean that the two proteins have similar amino acid sequences.) As used herein, homology between two regions of amino acid sequence (especially with respect to predicted structural similarities) is interpreted as implying similarity in function.

When "homologous" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of homology may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. See, e.g., Pearson, 1994, *Methods Mol. Biol.* 24:307-31 and 25:365-89 (herein incorporated by reference).

The following six groups each contain amino acids that are conservative substitutions for one another: 1) Serine (S), Threonine (T); 2) Aspartic Acid (D), Glutamic Acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Sequence homology for polypeptides, which is also referred to as percent sequence identity, is typically measured using sequence analysis software. See, e.g., the Sequence Analysis Software Package of the Genetics Computer Group (GCG), University of Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wis. 53705. Protein analysis software matches similar sequences using a measure of homology assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild-type protein and a mutin thereof. See, e.g., GCG Version 6.1.

A preferred algorithm when comparing a particular polypeptide sequence to a database containing a large number of sequences from different organisms is the computer program BLAST (Altschul et al., *J. Mol. Biol.* 215:403-410 (1990); Gish and States, *Nature Genet.* 3:266-272 (1993); Madden et al., *Meth. Enzymol.* 266:131-141 (1996); Altschul et al., *Nucleic Acids Res.* 25:3389-3402 (1997); Zhang and Madden, *Genome Res.* 7:649-656 (1997)), especially blastp or tblastn (Altschul et al., *Nucleic Acids Res.* 25:3389-3402 (1997)).

Preferred parameters for BLASTp are: Expectation value: 10 (default); Filter: seg (default); Cost to open a gap: 11 (default); Cost to extend a gap: 1 (default); Max. alignments: 100 (default); Word size: 11 (default); No. of descriptions: 100 (default); Penalty Matrix: BLOWSUM62.

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences. Database searching using amino acid sequences can be measured by algorithms other than blastp known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences. Pearson, *Methods Enzymol.* 183:63-98 (1990) (incorporated by reference herein). For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

"Specific binding" refers to the ability of two molecules to bind to each other in preference to binding to other molecules in the environment. Typically, "specific binding" discriminates over adventitious binding in a reaction by at least two-fold, more typically by at least 10-fold, often at least 100-fold. Typically, the affinity or avidity of a specific binding reaction, as quantified by a dissociation constant, is about 10^{-7} M or stronger (e.g., about 10^{-8} M, 10^{-9} M or even stronger).

"Percent dry cell weight" refers to a measurement of hydrocarbon production obtained as follows: a defined vol-

ume of culture is centrifuged to pellet the cells. Cells are washed then dewetted by at least one cycle of microcentrifugation and aspiration. Cell pellets are lyophilized overnight, and the tube containing the dry cell mass is weighed again such that the mass of the cell pellet can be calculated within ± 0.1 mg. At the same time cells are processed for dry cell weight determination, a second sample of the culture in question is harvested, washed, and dewetted. The resulting cell pellet, corresponding to 1-3 mg of dry cell weight, is then extracted by vortexing in approximately 1 ml acetone plus butylated hydroxytoluene (BHT) as antioxidant and an internal standard, e.g., n-heptacosane. Cell debris is then pelleted by centrifugation and the supernatant (extractant) is taken for analysis by GC. For accurate quantitation of n-alkanes, flame ionization detection (FID) was used as opposed to MS total ion count. n-Alkane concentrations in the biological extracts were calculated using calibration relationships between GC-FID peak area and known concentrations of authentic n-alkane standards. Knowing the volume of the extractant, the resulting concentrations of the n-alkane species in the extractant, and the dry cell weight of the cell pellet extracted, the percentage of dry cell weight that comprised n-alkanes can be determined.

The term "region" as used herein refers to a physically contiguous portion of the primary structure of a biomolecule. In the case of proteins, a region is defined by a contiguous portion of the amino acid sequence of that protein.

The term "domain" as used herein refers to a structure of a biomolecule that contributes to a known or suspected function of the biomolecule. Domains may be co-extensive with regions or portions thereof; domains may also include distinct, non-contiguous regions of a biomolecule. Examples of protein domains include, but are not limited to, an Ig domain, an extracellular domain, a transmembrane domain, and a cytoplasmic domain.

As used herein, the term "molecule" means any compound, including, but not limited to, a small molecule, peptide, protein, sugar, nucleotide, nucleic acid, lipid, etc., and such a compound can be natural or synthetic.

"Carbon-based Products of Interest" include alcohols such as ethanol, propanol, isopropanol, butanol, fatty alcohols, fatty acid esters, wax esters; hydrocarbons and alkanes such as propane, octane, diesel, Jet Propellant 8 (JP8); polymers such as terephthalate, 1,3-propanediol, 1,4-butanediol, polyols, Polyhydroxyalkanoates (PHA), poly-beta-hydroxybutyrate (PHB), acrylate, adipic acid, ϵ -caprolactone, isoprene, caprolactam, rubber; commodity chemicals such as lactate, Docosahexaenoic acid (DHA), 3-hydroxypropionate, γ -valerolactone, lysine, serine, aspartate, aspartic acid, sorbitol, ascorbate, ascorbic acid, isopentenol, lanosterol, omega-3 DHA, lycopene, itaconate, 1,3-butadiene, ethylene, propylene, succinate, citrate, citric acid, glutamate, malate, 3-hydroxypropionic acid (HPA), lactic acid, THF, gamma butyrolactone, pyrrolidones, hydroxybutyrate, glutamic acid, levulinic acid, acrylic acid, malonic acid; specialty chemicals such as carotenoids, isoprenoids, itaconic acid; pharmaceuticals and pharmaceutical intermediates such as 7-aminodeacetoxycephalosporanic acid (7-ADCA)/cephalosporin, erythromycin, polyketides, statins, paclitaxel, docetaxel, terpenes, peptides, steroids, omega fatty acids and other such suitable products of interest. Such products are useful in the context of biofuels, industrial and specialty chemicals, as intermediates used to make additional products, such as nutritional supplements, nutraceuticals, polymers, paraffin replacements, personal care products and pharmaceuticals.

Biofuel: A biofuel refers to any fuel that derives from a biological source. Biofuel can refer to one or more hydrocarbons, one or more alcohols, one or more fatty esters or a mixture thereof.

Hydrocarbon: The term generally refers to a chemical compound that consists of the elements carbon (C), hydrogen (H) and optionally oxygen (O). There are essentially three types of hydrocarbons, e.g., aromatic hydrocarbons, saturated hydrocarbons and unsaturated hydrocarbons such as alkenes, alkynes, and dienes. The term also includes fuels, biofuels, plastics, waxes, solvents and oils. Hydrocarbons encompass biofuels, as well as plastics, waxes, solvents and oils.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this present invention pertains. Exemplary methods and materials are described below, although methods and materials similar or equivalent to those described herein can also be used in the practice of the present invention and will be apparent to those of skill in the art. All publications and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. The materials, methods, and examples are illustrative only and not intended to be limiting.

Throughout this specification and claims, the word "comprise" or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

Nucleic Acid Sequences

Alkanes, also known as paraffins, are chemical compounds that consist only of the elements carbon (C) and hydrogen (H) (i.e., hydrocarbons), wherein these atoms are linked together exclusively by single bonds (i.e., they are saturated compounds) without any cyclic structure. n-Alkanes are linear, i.e., unbranched, alkanes. Together, AAR and ADM enzymes function to synthesize n-alkanes from acyl-ACP molecules.

Accordingly, the present invention provides isolated nucleic acid molecules for genes encoding AAR and ADM enzymes, and variants thereof. Exemplary full-length nucleic acid sequences for genes encoding AAR are presented as SEQ ID NOs: 1, 5, and 13, and the corresponding amino acid sequences are presented as SEQ ID NOs: 2, 6, and 10, respectively. Exemplary full-length nucleic acid sequences for genes encoding ADM are presented as SEQ ID NOs: 3, 7, 14, and the corresponding amino acid sequences are presented as SEQ ID NOs: 4, 8, and 12, respectively. Additional nucleic acids provided by the invention include any of the genes encoding the AAR and ADM enzymes in Table 1 and Table 2, respectively.

In one embodiment, the present invention provides an isolated nucleic acid molecule having a nucleic acid sequence comprising or consisting of a gene coding for AAR and ADM, and homologs, variants and derivatives thereof expressed in a host cell of interest. The present invention also provides a nucleic acid molecule comprising or consisting of a sequence which is a codon-optimized version of the AAR and ADM genes described herein (e.g., SEQ ID NO: 9 and SEQ ID NO: 11, which are optimized for the expression of the AAR and ADM genes of *Prochlorococcus marinus* MIT 9312 in *Synechococcus* sp. PCC 7002). In a further embodiment, the present invention provides a nucleic acid molecule and homologs, variants and derivatives of the molecule comprising or consisting of a sequence which is a variant of the AAR or ADM gene having at least 76% identity to the wild-type

gene. The nucleic acid sequence can be preferably 80%, 85%, 90%, 95%, 98%, 99%, 99.9% or even higher identity to the wild-type gene.

In another embodiment, the nucleic acid molecule of the present invention encodes a polypeptide having the amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10 or 12. Preferably, the nucleic acid molecule of the present invention encodes a polypeptide sequence of at least 50%, 60, 70%, 80%, 85%, 90% or 95% identity to SEQ ID NO: 2, 4, 6, 8, 10 or 12 and the identity can even more preferably be 96%, 97%, 98%, 99%, 99.9% or even higher.

The present invention also provides nucleic acid molecules that hybridize under stringent conditions to the above-described nucleic acid molecules. As defined above, and as is well known in the art, stringent hybridizations are performed at about 25° C. below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions, where the T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. Stringent washing is performed at temperatures about 5° C. lower than the T_m for the specific DNA hybrid under a particular set of conditions.

Nucleic acid molecules comprising a fragment of any one of the above-described nucleic acid sequences are also provided. These fragments preferably contain at least 20 contiguous nucleotides. More preferably the fragments of the nucleic acid sequences contain at least 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or even more contiguous nucleotides.

The nucleic acid sequence fragments of the present invention display utility in a variety of systems and methods. For example, the fragments may be used as probes in various hybridization techniques. Depending on the method, the target nucleic acid sequences may be either DNA or RNA. The target nucleic acid sequences may be fractionated (e.g., by gel electrophoresis) prior to the hybridization, or the hybridization may be performed on samples in situ. One of skill in the art will appreciate that nucleic acid probes of known sequence find utility in determining chromosomal structure (e.g., by Southern blotting) and in measuring gene expression (e.g., by Northern blotting). In such experiments, the sequence fragments are preferably detectably labeled, so that their specific hybridization to target sequences can be detected and optionally quantified. One of skill in the art will appreciate that the nucleic acid fragments of the present invention may be used in a wide variety of blotting techniques not specifically described herein.

It should also be appreciated that the nucleic acid sequence fragments disclosed herein also find utility as probes when immobilized on microarrays. Methods for creating microarrays by deposition and fixation of nucleic acids onto support substrates are well known in the art. Reviewed in *DNA Microarrays: A Practical Approach* (Practical Approach Series), Schena (ed.), Oxford University Press (1999) (ISBN: 0199637768); *Nature Genet.* 21(1)(suppl):1-60 (1999); *Microarray Biochip: Tools and Technology*, Schena (ed.), Eaton Publishing Company/BioTechniques Books Division (2000) (ISBN: 1881299376), the disclosures of which are incorporated herein by reference in their entireties. Analysis of, for example, gene expression using microarrays comprising nucleic acid sequence fragments, such as the nucleic acid sequence fragments disclosed herein, is a well-established utility for sequence fragments in the field of cell and molecular biology. Other uses for sequence fragments immobilized on microarrays are described in Gerhold et al., *Trends Biochem. Sci.* 24:168-173 (1999) and Zweiger, *Trends Biotechnol.* 17:429-436 (1999); *DNA Microarrays: A Practical Approach* (Practical Approach Series), Schena (ed.), Oxford

University Press (1999) (ISBN: 0199637768); Nature Genet. 21(1)(suppl):1-60 (1999); *Microarray Biochip: Tools and Technology*, Schena (ed.), Eaton Publishing Company/Bio-Techniques Books Division (2000) (ISBN: 1881299376), the disclosure of each of which is incorporated herein by reference in its entirety.

As is well known in the art, enzyme activities can be measured in various ways. For example, the pyrophosphorylation of OMP may be followed spectroscopically (Grubmeyer et al., (1993) *J. Biol. Chem.* 268:20299-20304). Alternatively, the activity of the enzyme can be followed using chromatographic techniques, such as by high performance liquid chromatography (Chung and Sloan, (1986) *J. Chromatogr.* 371:71-81). As another alternative the activity can be indirectly measured by determining the levels of product made from the enzyme activity. These levels can be measured with techniques including aqueous chloroform/methanol extraction as known and described in the art (Cf M. Kates (1986) *Techniques of Lipidology; Isolation, analysis and identification of Lipids*. Elsevier Science Publishers, New York (ISBN: 0444807322)). More modern techniques include using gas chromatography linked to mass spectrometry (Niessen, W. M. A. (2001). *Current practice of gas chromatography—mass spectrometry*. New York, N.Y: Marcel Dekker. (ISBN: 0824704738)). Additional modern techniques for identification of recombinant protein activity and products including liquid chromatography-mass spectrometry (LCMS), high performance liquid chromatography (HPLC), capillary electrophoresis, Matrix-Assisted Laser Desorption Ionization time of flight-mass spectrometry (MALDI-TOF MS), nuclear magnetic resonance (NMR), near-infrared (NIR) spectroscopy, viscometry (Knothe, G (1997) *Am. Chem. Soc. Symp. Series*, 666: 172-208), titration for determining free fatty acids (Komers (1997) *Fett/Lipid*, 99(2): 52-54), enzymatic methods (Bailer (1991) *Fresenius J. Anal. Chem.* 340(3): 186), physical property-based methods, wet chemical methods, etc. can be used to analyze the levels and the identity of the product produced by the organisms of the present invention. Other methods and techniques may also be suitable for the measurement of enzyme activity, as would be known by one of skill in the art.

Vectors

Also provided are vectors, including expression vectors, which comprise the above nucleic acid molecules of the present invention, as described further herein. In a first embodiment, the vectors include the isolated nucleic acid molecules described above. In an alternative embodiment, the vectors of the present invention include the above-described nucleic acid molecules operably linked to one or more expression control sequences. The vectors of the instant invention may thus be used to express an AAR and/or ADM polypeptide contributing to n-alkane producing activity by a host cell.

Vectors useful for expression of nucleic acids in prokaryotes are well known in the art.

Isolated Polypeptides

According to another aspect of the present invention, isolated polypeptides (including muteins, allelic variants, fragments, derivatives, and analogs) encoded by the nucleic acid molecules of the present invention are provided. In one embodiment, the isolated polypeptide comprises the polypeptide sequence corresponding to SEQ ID NO:2, 4, 6, 8, 10 or 12. In an alternative embodiment of the present invention, the isolated polypeptide comprises a polypeptide sequence at least 85% identical to SEQ ID NO:2, 4, 6, 8, 10 or 12. Preferably the isolated polypeptide of the present invention has at least 50%, 60, 70%, 80%, 85%, 90%, 95%, 98%, 98.1%, 98.2%, 98.3%, 98.4%, 98.5%, 98.6%, 98.7%, 98.8%,

98.9%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or even higher identity to SEQ ID NO:2, 4, 6, 8, 10 or 12.

According to other embodiments of the present invention, isolated polypeptides comprising a fragment of the above-described polypeptide sequences are provided. These fragments preferably include at least 20 contiguous amino acids, more preferably at least 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or even more contiguous amino acids.

The polypeptides of the present invention also include fusions between the above-described polypeptide sequences and heterologous polypeptides. The heterologous sequences can, for example, include sequences designed to facilitate purification, e.g. histidine tags, and/or visualization of recombinantly-expressed proteins. Other non-limiting examples of protein fusions include those that permit display of the encoded protein on the surface of a phage or a cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region.

Host Cell Transformants

In another aspect of the present invention, host cells transformed with the nucleic acid molecules or vectors of the present invention, and descendants thereof, are provided. In some embodiments of the present invention, these cells carry the nucleic acid sequences of the present invention on vectors, which may but need not be freely replicating vectors. In other embodiments of the present invention, the nucleic acids have been integrated into the genome of the host cells.

In a preferred embodiment, the host cell comprises one or more AAR or ADM encoding nucleic acids which express AAR or ADM in the host cell.

In an alternative embodiment, the host cells of the present invention can be mutated by recombination with a disruption, deletion or mutation of the isolated nucleic acid of the present invention so that the activity of the AAR and/or ADM protein(s) in the host cell is reduced or eliminated compared to a host cell lacking the mutation.

Selected or Engineered Microorganisms for the Production of Carbon-Based Products of Interest

Microorganism: Includes prokaryotic and eukaryotic microbial species from the Domains Archaea, Bacteria and Eucarya, the latter including yeast and filamentous fungi, protozoa, algae, or higher Protista. The terms “microbial cells” and “microbes” are used interchangeably with the term microorganism.

A variety of host organisms can be transformed to produce a product of interest. Photoautotrophic organisms include eukaryotic plants and algae, as well as prokaryotic cyanobacteria, green-sulfur bacteria, green non-sulfur bacteria, purple sulfur bacteria, and purple non-sulfur bacteria.

Extremophiles are also contemplated as suitable organisms. Such organisms withstand various environmental parameters such as temperature, radiation, pressure, gravity, vacuum, desiccation, salinity, pH, oxygen tension, and chemicals. They include hyperthermophiles, which grow at or above 80° C. such as *Pyrolobus fumarii*; thermophiles, which grow between 60-80° C. such as *Synechococcus lividis*; mesophiles, which grow between 15-60° C. and psychrophiles, which grow at or below 15° C. such as *Psychrobacter* and some insects. Radiation tolerant organisms include *Deinococcus radiodurans*. Pressure-tolerant organisms include piezophiles, which tolerate pressure of 130 MPa. Weight-tolerant organisms include barophiles. Hypergravity (e.g., >1 g) hypogravity (e.g., <1 g) tolerant organisms are also contemplated. Vacuum tolerant organisms include tardigrades, insects, microbes and seeds. Dessiccant tolerant and anhydrobiotic organisms include xerophiles such as *Artemia*

salina; nematodes, microbes, fungi and lichens. Salt-tolerant organisms include halophiles (e.g., 2-5 M NaCl) *Halobacteriaceae* and *Dunaliella salina*. pH-tolerant organisms include alkaliphiles such as *Natronobacterium*, *Bacillus firmus* OF4, *Spirulina* spp. (e.g., pH >9) and acidophiles such as *Cyanidium caldarium*, *Ferroplasma* sp. (e.g., low pH). Anaerobes, which cannot tolerate O₂ such as *Methanococcus jannaschii*; microaerophils, which tolerate some O₂ such as *Clostridium* and aerobes, which require O₂ are also contemplated. Gas-tolerant organisms, which tolerate pure CO₂ include *Cyanidium caldarium* and metal tolerant organisms include metalotolerants such as *Ferroplasma acidarmanus* (e.g., Cu, As, Cd, Zn), *Ralstonia* sp. CH34 (e.g., Zn, Co, Cd, Hg, Pb). Gross, Michael. *Life on the Edge: Amazing Creatures Thriving in Extreme Environments*. New York: Plenum (1998) and Seckbach, J. "Search for Life in the Universe with Terrestrial Microbes Which Thrive Under Extreme Conditions." In Cristiano Batalli Cosmovici, Stuart Bowyer, and Dan Wertheimer, eds., *Astronomical and Biochemical Origins and the Search for Life in the Universe*, p. 511. Milan: Editrice Compositori (1997).

Plants include but are not limited to the following genera: *Arabidopsis*, *Beta*, *Glycine*, *Jatropha*, *Miscanthus*, *Panicum*, *Phalaris*, *Populus*, *Saccharum*, *Salix*, *Simmondsia* and *Zea*.

Algae and cyanobacteria include but are not limited to the following genera: *Acanthoceras*, *Acanthococcus*, *Acaryochloris*, *Achnanthes*, *Achnantheidium*, *Actinastrum*, *Actinochloris*, *Actinocyclus*, *Actinotaenium*, *Amphichrysis*, *Amphidinium*, *Amphikrikos*, *Amphipleura*, *Amphiprora*, *Amphithrix*, *Amphora*, *Anabaena*, *Anabaenopsis*, *Aneomastus*, *Ankistrodesmus*, *Ankyra*, *Anomooneis*, *Apatococcus*, *Aphanizomenon*, *Aphanocapsa*, *Aphanochaete*, *Aphanothece*, *Apiocystis*, *Apistonema*, *Arthrodesmus*, *Artherospira*, *Ascochloris*, *Asterionella*, *Asterococcus*, *Audouinella*, *Aulacoseira*, *Bacillaria*, *Balbiana*, *Bambusina*, *Bangia*, *Basichlamys*, *Batrachospermum*, *Binuclearia*, *Bitrichia*, *Bliedingia*, *Botrydiopsis*, *Botrydium*, *Botryococcus*, *Botryosphaerella*, *Brachiomonas*, *Brachysira*, *Brachytrichia*, *Brebissonia*, *Bulbochaete*, *Bumilleria*, *Bumilleriopsis*, *Caloneis*, *Calothrix*, *Campylodiscus*, *Capsosiphon*, *Carteria*, *Catena*, *Cavinula*, *Centritractus*, *Centronella*, *Ceratium*, *Chaetoceros*, *Chaetochloris*, *Chaetomorpha*, *Chaetonella*, *Chaetonema*, *Chaetopeltis*, *Chaetophora*, *Chaetosphaeridium*, *Chamaesiphon*, *Chara*, *Characiochloris*, *Characiopsis*, *Characium*, *Charales*, *Chilomonas*, *Chlainomonas*, *Chlamydolepharis*, *Chlamydocapsa*, *Chlamydomonas*, *Chlamydomonopsis*, *Chlamydomyxa*, *Chlamydonephris*, *Chlorangiella*, *Chlorangiopsis*, *Chlorella*, *Chlorobotrys*, *Chlorobranchis*, *Chlorochytrium*, *Chlorococcum*, *Chlorogloea*, *Chlorogloeopsis*, *Chlorogonium*, *Chlorolobion*, *Chloromonas*, *Chlorophysemata*, *Chlorophyta*, *Chlorosaccus*, *Chlorosarcina*, *Choricystis*, *Chromophyton*, *Chromulina*, *Chroococcidiopsis*, *Chroococcus*, *Chroodactylon*, *Chroomonas*, *Chroothece*, *Chrysamoeba*, *Chrysapsis*, *Chrysidiastrum*, *Chrysocapsa*, *Chrysocapsella*, *Chrysochaete*, *Chrysochromulina*, *Chrysococcus*, *Chrysocrinus*, *Chrysolepidomonas*, *Chrysolykos*, *Chrysonebula*, *Chrysoophyta*, *Chrysopyxis*, *Chrysosaccus*, *Chrysosphaerella*, *Chrysostephanosphaera*, *Clodophora*, *Clastidium*, *Closteriopsis*, *Closterium*, *Coccomyxa*, *Cocconeis*, *Coelastrella*, *Coelastrium*, *Coelosphaerium*, *Coenochloris*, *Coenococcus*, *Coenocystis*, *Colacium*, *Coleochaete*, *Collodictyon*, *Composgonopsis*, *Composopogon*, *Conjugatophyta*, *Conochaete*, *Coronastrum*, *Cosmarium*, *Cosmioneis*, *Cosmoctadium*, *Crateriportula*, *Craticula*, *Crinalium*, *Crucigenia*, *Crucigeniella*, *Cryptoaulax*, *Cryptomonas*, *Cryptophyta*, *Ctenophora*, *Cyanodictyon*, *Cyanonephron*, *Cyanophora*, *Cyano-*

phyta, *Cyanothece*, *Cyanothomonas*, *Cyclonexis*, *Cyclostephanos*, *Cyclotella*, *Cylindrocapsa*, *Cylindrocystis*, *Cylindrospermum*, *Cylindrotheca*, *Cymatopleura*, *Cymbella*, *Cymbellonitzschia*, *Cystodinium*, *Dactylococcopsis*, *Debarya*, *Denticula*, *Dermatochrysis*, *Dermocarpa*, *Dermocarpella*, *Desmatractum*, *Desmidium*, *Desmococcus*, *Desmonema*, *Desmosiphon*, *Diacanthos*, *Diacronema*, *Diademis*, *Diatoma*, *Diatomella*, *Dicellula*, *Dichothrix*, *Dichotomococcus*, *Dicranochaete*, *Dictyochloris*, *Dictyococcus*, *Dictyosphaerium*, *Didymocystis*, *Didymogenes*, *Didymosphenia*, *Dilabifilum*, *Dimorphococcus*, *Dinobryon*, *Dinococcus*, *Diplochloris*, *Diploneis*, *Diplostauron*, *Distriponella*, *Docidium*, *Draparnaldia*, *Dunaliella*, *Dysmorphococcus*, *Ecballocystis*, *Elakatothrix*, *Ellerbeckia*, *Encyonema*, *Enteromorpha*, *Entocladia*, *Entomoneis*, *Entophysalis*, *Epicchrysis*, *Epipyxis*, *Epithemia*, *Eremosphaera*, *Euastrum*, *Euastrum*, *Eucapsis*, *Eucoconneis*, *Eudorina*, *Euglena*, *Euglenophyta*, *Eunotia*, *Eustigmatophyta*, *Eutreptia*, *Fallacia*, *Fischerella*, *Fragilaria*, *Fragilariforma*, *Franceia*, *Frustulia*, *Curcilla*, *Geminella*, *Genicularia*, *Glaucocystis*, *Glaucophyta*, *Glenodiniopsis*, *Glenodinium*, *Gloeocapsa*, *Gloeochaete*, *Gloeochrysis*, *Gloeococcus*, *Gloeocystis*, *Gloeodendron*, *Gloeomonas*, *Gloeoplax*, *Gloeothece*, *Gloeotila*, *Gloeotrichia*, *Gloiodictyon*, *Golenkinia*, *Golenkiniopsis*, *Gomontia*, *Gomphocymbella*, *Gomphonema*, *Gomphosphaeria*, *Gonatozygon*, *Gongrosia*, *Gongrosira*, *Goniocchloris*, *Gonium*, *Gonyostomum*, *Granulochloris*, *Granulocystopsis*, *Groenbladia*, *Gymnodinium*, *Gymnozyga*, *Gyrosigma*, *Haematococcus*, *Hafniomonas*, *Hallassia*, *Hammatoidea*, *Hannaea*, *Hantzschia*, *Hapalosiphon*, *Haplotaeonium*, *Haptophyta*, *Haslea*, *Hemidinium*, *Hemitoma*, *Heribaudiella*, *Heteromastix*, *Heterothrix*, *Hibberdia*, *Hildenbrandia*, *Hillea*, *Holopedium*, *Homoeothrix*, *Hormanthonema*, *Hormotila*, *Hyalobranchion*, *Hyalocardium*, *Hyalodiscus*, *Hyalogonium*, *Hyalotheca*, *Hydranium*, *Hydrococcus*, *Hydrocoleum*, *Hydrocoryne*, *Hydrodictyon*, *Hydrosera*, *Hydrurus*, *Hyella*, *Hymenomonas*, *Isthmochloron*, *Johannesbaptistia*, *Juranyiella*, *Karayevia*, *Kathablepharis*, *Katodinium*, *Kephyrion*, *Keratococcus*, *Kirchneriella*, *Klebsormidium*, *Kolbesia*, *Koliella*, *Komarekia*, *Korshikoviella*, *Kraskella*, *Lagerheimia*, *Lagynion*, *Lamprothamnium*, *Lemanea*, *Lepocinclis*, *Leptosira*, *Lobococcus*, *Lobocystis*, *Lobomonas*, *Luticola*, *Lyngbya*, *Malleochloris*, *Mallomonas*, *Mantoniella*, *Marssoniella*, *Martyana*, *Mastigocoleus*, *Gastogloia*, *Melosira*, *Merismopedia*, *Mesostigma*, *Mesotaeonium*, *Microactinium*, *Microasterias*, *Microchaete*, *Microcoleus*, *Microcystis*, *Microglana*, *Micromonas*, *Microspora*, *Microthamnion*, *Mischococcus*, *Monochrysis*, *Monodus*, *Monomastix*, *Monoraphidium*, *Monostroma*, *Mougeotia*, *Mougeotiopsis*, *Myochloris*, *Myromecia*, *Myxosarcina*, *Naegelella*, *Nannochloris*, *Nautococcus*, *Navicula*, *Neglectella*, *Neidium*, *Nephroclamys*, *Nephrocycium*, *Nephrodiella*, *Nephroselmis*, *Netrium*, *Nitella*, *Nitellopsis*, *Nitzschia*, *Nodularia*, *Nostoc*, *Ochromonas*, *Oedogonium*, *Oligochaetophora*, *Onychonema*, *Oocardium*, *Oocystis*, *Opephora*, *Ophiocytium*, *Orthoseira*, *Oscillatoria*, *Oxyneis*, *Pachycladella*, *Palmella*, *Palmodictyon*, *Pnadorina*, *Pannus*, *Paralia*, *Pascherina*, *Paulschulzia*, *Pediastrum*, *Pedinella*, *Pedinomonas*, *Pedinopera*, *Pelagodictyon*, *Penium*, *Peranema*, *Peridiniopsis*, *Peridinium*, *Peronia*, *Petroneis*, *Phacotus*, *Phacus*, *Phaeaster*, *Phaeodermatium*, *Phaeophyta*, *Phaeosphaera*, *Phaeothamnion*, *Phormidium*, *Phycopeltis*, *Phyllariochloris*, *Phyllocardium*, *Phyllomita*, *Pinnularia*, *Pitophora*, *Placoneis*, *Planctonema*, *Planktosphaeria*, *Planothidium*, *Plectonema*, *Pleodorina*, *Pleurastrum*, *Pleurocapsa*, *Pleurocladia*, *Pleurodiscus*, *Pleurosigma*, *Pleurosira*, *Pleurotaenium*, *Pocillomonas*, *Podohedra*, *Polyblepharides*,

Polychaetophora, *Polyedriella*, *Polyedriopsis*, *Polygoniochloris*, *Polyepidomonas*, *Polytaenia*, *Polytoma*, *Polytomella*, *Porphyridium*, *Posterochromonas*, *Prasinocloris*, *Prasinocladus*, *Prasinophyta*, *Prasiola*, *Prochlorophyta*, *Prochlorothrix*, *Protoderma*, *Protosiphon*, *Provasoliella*, *Prymnesium*, *Psammodictyon*, *Psammothidium*, *Pseudanabaena*, *Pseudonoclonium*, *Pseudocarteria*, *Pseudochate*, *Pseudocharacium*, *Pseudococcomyxa*, *Pseudodictyosphaerium*, *Pseudokephyrion*, *Pseudoncohyrsa*, *Pseudoquadrigula*, *Pseudosphaerocystis*, *Pseudostaurastrum*, *Pseudostaurisira*, *Pseudotetrastrum*, *Pteromonas*, *Punctastruata*, *Pyramichlamys*, *Pyramimonas*, *Pyrrophyta*, *Quadrichloris*, *Quadrircoccus*, *Quadrigula*, *Radiococcus*, *Radiofilum*, *Raphidiopsis*, *Raphidocelis*, *Raphidonema*, *Raphidophyta*, *Peimeria*, *Rhabdoderma*, *Rhabdomonas*, *Rhizoclonium*, *Rhodomonas*, *Rhodophyta*, *Rhoicosphenia*, *Rhopalodia*, *Rivularia*, *Rosenvingiella*, *Rossithidium*, *Roya*, *Scenedesmus*, *Scherffelia*, *Schizochlamydeia*, *Schizochlamys*, *Schizomeris*, *Schizothrix*, *Schroederia*, *Scolioneis*, *Scotiella*, *Scotiellopsis*, *Scourfieldia*, *Scytonema*, *Selenastrum*, *Selenochloris*, *Sellaphora*, *Semiorbis*, *Siderocelis*, *Diderocystopsis*, *Dimonsenia*, *Siphononema*, *Sirocladium*, *Sirogonium*, *Skeletonema*, *Sorastrum*, *Spermatozopsis*, *Sphaerello cystis*, *Sphaerellopsis*, *Sphaerodinium*, *Sphaeroplea*, *Sphaerozosma*, *Spiniferomonas*, *Spirogyra*, *Spirotaenia*, *Spirulina*, *Spondylomorom*, *Spondylosium*, *Sporotetras*, *Spumella*, *Staurastrum*, *Stauerodesmus*, *Stauroneis*, *Staurosira*, *Staurosirella*, *Stenopterobia*, *Stephanocostis*, *Stephanodiscus*, *Stephanoporos*, *Stephanosphaera*, *Stichococcus*, *Stichogloea*, *Stigeoclonium*, *Stigonema*, *Stipitococcus*, *Stokesiella*, *Strombomonas*, *Stylochrysalis*, *Stylocladus*, *Styloxyis*, *Stylosphaeridium*, *Surirella*, *Sykidion*, *Symploca*, *Synechococcus*, *Synechocystis*, *Synedra*, *Synochromonas*, *Synura*, *Tabellaria*, *Tabularia*, *Teilingia*, *Temnogametum*, *Tetmemorus*, *Tetrachlorella*, *Tetracyclus*, *Tetrademus*, *Tetraedriella*, *Tetraedron*, *Tetraselmis*, *Tetraspora*, *Tetrastrum*, *Thalassiosira*, *Thamniochaete*, *Thorakochloris*, *Thorea*, *Tolypella*, *Tolypothrix*, *Trachelomonas*, *Trachydiscus*, *Trebouxia*, *Trentepohlia*, *Treubaria*, *Tribonema*, *Trichodesmium*, *Trichodiscus*, *Trochiscia*, *Tryblionella*, *Ulothrix*, *Uroglana*, *Uronema*, *Urosolenia*, *Urospora*, *Uva*, *Vacuolaria*, *Vaucheria*, *Volvox*, *Volvulina*, *Westella*, *Woloszynskia*, *Xanthidium*, *Xanthophyta*, *Xenococcus*, *Zygnema*, *Zygnemopsis*, and *Zygonium*. A partial list of cyanobacteria that can be engineered to express recombinant AAR and ADM enzymes is also provided in Table 1 and Table 2, herein. Additional cyanobacteria include members of the genus *Chamaesiphon*, *Chroococcus*, *Cyanobacterium*, *Cyanobium*, *Cyanothece*, *Dactylococcopsis*, *Gloeobacter*, *Gloeocapsa*, *Gloeothece*, *Microcystis*, *Prochlorococcus*, *Prochloron*, *Synechococcus*, *Synechocystis*, *Cyanocystis*, *Dermocarpella*, *Stanieria*, *Xenococcus*, *Chroococcidiopsis*, *Myxosarcina*, *Arthrospira*, *Borzia*, *Criminalium*, *Geitlerinemia*, *Leptolyngbya*, *Limnothrix*, *Lyngbya*, *Microcoleus*, *Oscillatoria*, *Planktothrix*, *Prochlorothrix*, *Pseudanabaena*, *Spirulina*, *Starria*, *Symploca*, *Trichodesmium*, *Tychonema*, *Anabaena*, *Anabaenopsis*, *Aphanizomenon*, *Cyanospira*, *Cylindrospermopsis*, *Cylindrospermum*, *Nodularia*, *Nostoc*, *Scytonema*, *Calothrix*, *Rivularia*, *Tolypothrix*, *Chlorogloeopsis*, *Fischerella*, *Geitleria*, *Iyengariella*, *Nostochopsis*, *Stigonema* and *Thermosynechococcus*.

Green non-sulfur bacteria include but are not limited to the following genera: *Chloroflexus*, *Chloronema*, *Oscillochloris*, *Heliothrix*, *Herpetosiphon*, *Roseiflexus*, and *Thermomicrobium*.

Green sulfur bacteria include but are not limited to the following genera:

Chlorobium, *Clathrochloris*, and *Prosthecochloris*.

Purple sulfur bacteria include but are not limited to the following genera: *Allochromatium*, *Chromatium*, *Halochromatium*, *Isochromatium*, *Marichromatium*, *Rhodovulum*, *Thermochromatium*, *Thiocapsa*, *Thiorhodococcus*, and *Thiocystis*,

Purple non-sulfur bacteria include but are not limited to the following genera: *Phaeospirillum*, *Rhodobaca*, *Rhodobacter*, *Rhodomicrobium*, *Rhodopila*, *Rhodopseudomonas*, *Rhodotalassium*, *Rhodospirillum*, *Rhodovibrio*, and *Roseospira*.

Aerobic chemolithotrophic bacteria include but are not limited to nitrifying bacteria such as *Nitrobacteraceae* sp., *Nitrobacter* sp., *Nitrospina* sp., *Nitrococcus* sp., *Nitrospira* sp., *Nitrosomonas* sp., *Nitrosococcus* sp., *Nitrosospira* sp., *Nitrosolobus* sp., *Nitrosovibrio* sp.; colorless sulfur bacteria such as *Thiovulum* sp., *Thiobacillus* sp., *Thiomicrospira* sp., *Thiosphaera* sp., *Thermothrix* sp.; obligately chemolithotrophic hydrogen bacteria such as *Hydrogenobacter* sp., iron and manganese-oxidizing and/or depositing bacteria such as *Siderococcus* sp., and magnetotactic bacteria such as *Aquaspirillum* sp.

Archaeobacteria include but are not limited to methanogenic archaeobacteria such as *Methanobacterium* sp., *Methanobrevibacter* sp., *Methanothermobacter* sp., *Methanococcus* sp., *Methanomicrobium* sp., *Methanospirillum* sp., *Methanogenium* sp., *Methanosarcina* sp., *Methanobolus* sp., *Methanothrix* sp., *Methanococcoides* sp., *Methanoplanus* sp.; extremely thermophilic S-Metabolizers such as *Thermoproteus* sp., *Pyrodictium* sp., *Sulfolobus* sp., *Acidianus* sp. and other microorganisms such as, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Streptomyces* sp., *Ralstonia* sp., *Rhodococcus* sp., *Corynebacteria* sp., *Brevibacteria* sp., *Mycobacteria* sp., and oleaginous yeast.

Preferred organisms for the manufacture of n-alkanes according to the methods disclosed herein include: *Arabidopsis thaliana*, *Panicum virgatum*, *Miscanthus giganteus*, and *Zea mays* (plants); *Botryococcus braunii*, *Chlamydomonas reinhardtii* and *Dunaliella salina* (algae); *Synechococcus* sp. PCC 7002, *Synechococcus* sp. PCC 7942, *Synechocystis* sp. PCC 6803, *Thermosynechococcus elongatus* BP-1 (cyanobacteria); *Chlorobium tepidum* (green sulfur bacteria), *Chloroflexus auranticus* (green non-sulfur bacteria); *Chromatium* *rhodospirillum* and *Chromatium vinosum* (purple sulfur bacteria); *Rhodospirillum rubrum*, *Rhodobacter capsulatus*, and *Rhodopseudomonas palustris* (purple non-sulfur bacteria).

Yet other suitable organisms include synthetic cells or cells produced by synthetic genomes as described in Venter et al. US Pat. Pub. No. 2007/0264688, and cell-like systems or synthetic cells as described in Glass et al. US Pat. Pub. No. 2007/0269862.

Still, other suitable organisms include microorganisms that can be engineered to fix carbon dioxide bacteria such as *Escherichia coli*, *Acetobacter acetii*, *Bacillus subtilis*, yeast and fungi such as *Clostridium ljungdahlii*, *Clostridium thermocellum*, *Penicillium chrysogenum*, *Pichia pastoris*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pseudomonas fluorescens*, or *Zymomonas mobilis*.

A suitable organism for selecting or engineering is autotrophic fixation of CO₂ to products. This would cover photosynthesis and methanogenesis. Acetogenesis, encompassing the three types of CO₂ fixation; Calvin cycle, acetyl-CoA pathway and reductive TCA pathway is also covered. The capability to use carbon dioxide as the sole source of cell carbon (autotrophy) is found in almost all major groups of prokaryotes. The CO₂ fixation pathways differ between

groups, and there is no clear distribution pattern of the four presently-known autotrophic pathways. See, e.g., Fuchs, G. 1989. *Alternative pathways of autotrophic CO₂ fixation*, p. 365-382. In H. G. Schlegel, and B. Bowien (ed.), *Autotrophic bacteria*. Springer-Verlag, Berlin, Germany. The reductive pentose phosphate cycle (Calvin-Bassham-Benson cycle) represents the CO₂ fixation pathway in almost all aerobic autotrophic bacteria, for example, the cyanobacteria.

For producing n-alkanes via the recombinant expression of AAR and/or ADM enzymes, an engineered cyanobacteria, e.g., a *Synechococcus* or *Thermosynechococcus* species, is preferred. Other preferred organisms include *Synechocystis*, *Klebsiella oxytoca*, *Escherichia coli* or *Saccharomyces cerevisiae*. Other prokaryotic, archaea and eukaryotic host cells are also encompassed within the scope of the present invention.

Carbon-Based Products of Interest: Hydrocarbons & Alcohols

In various embodiments of the invention, desired hydrocarbons and/or alcohols of certain chain length or a mixture thereof can be produced. In certain aspects, the host cell produces at least one of the following carbon-based products of interest: 1-dodecanol, 1-tetradecanol, 1-pentadecanol, n-tridecane, n-tetradecane, 15:1 n-pentadecane, n-pentadecane, 16:1 n-hexadecene, n-hexadecane, 17:1 n-heptadecene, n-heptadecane, 16:1 n-hexadecen-ol, n-hexadecan-1-ol and n-octadecen-1-ol, as shown in the Examples herein. In other aspects, the carbon chain length ranges from C₁₀ to C₂₀. Accordingly, the invention provides production of various chain lengths of alkanes, alkenes and alkanols suitable for use as fuels & chemicals.

In preferred aspects, the methods provide culturing host cells for direct product secretion for easy recovery without the need to extract biomass. These carbon-based products of interest are secreted directly into the medium. Since the invention enables production of various defined chain length of hydrocarbons and alcohols, the secreted products are easily recovered or separated. The products of the invention, therefore, can be used directly or used with minimal processing.

Fuel Compositions

In various embodiments, compositions produced by the methods of the invention are used as fuels. Such fuels comply with ASTM standards, for instance, standard specifications for diesel fuel oils D 975-09b, and Jet A, Jet A-1 and Jet B as specified in ASTM Specification D. 1655-68. Fuel compositions may require blending of several products to produce a uniform product. The blending process is relatively straightforward, but the determination of the amount of each component to include in a blend is much more difficult. Fuel compositions may, therefore, include aromatic and/or branched hydrocarbons, for instance, 75% saturated and 25% aromatic, wherein some of the saturated hydrocarbons are branched and some are cyclic. Preferably, the methods of the invention produce an array of hydrocarbons, such as C₁₃-C₁₇ or C₁₀-C₁₅ to alter cloud point. Furthermore, the compositions may comprise fuel additives, which are used to enhance the performance of a fuel or engine. For example, fuel additives can be used to alter the freezing/gelling point, cloud point, lubricity, viscosity, oxidative stability, ignition quality, octane level, and flash point. Fuels compositions may also comprise, among others, antioxidants, static dissipater, corrosion inhibitor, icing inhibitor, biocide, metal deactivator and thermal stability improver.

In addition to many environmental advantages of the invention such as CO₂ conversion and renewable source, other advantages of the fuel compositions disclosed herein

include low sulfur content, low emissions, being free or substantially free of alcohol and having high cetane number. Carbon Fingerprinting

Biologically-produced carbon-based products, e.g., ethanol, fatty acids, alkanes, isoprenoids, represent a new commodity for fuels, such as alcohols, diesel and gasoline. Such biofuels have not been produced using biomass but use CO₂ as its carbon source. These new fuels may be distinguishable from fuels derived from petrochemical carbon on the basis of dual carbon-isotopic fingerprinting. Such products, derivatives, and mixtures thereof may be completely distinguished from their petrochemical derived counterparts on the basis of ¹⁴C (fM) and dual carbon-isotopic fingerprinting, indicating new compositions of matter.

There are three naturally occurring isotopes of carbon: ¹²C, ¹³C, and ¹⁴C. These isotopes occur in above-ground total carbon at fractions of 0.989, 0.011, and 10⁻¹², respectively. The isotopes ¹²C and ¹³C are stable, while ¹⁴C decays naturally to ¹⁴N, a beta particle, and an anti-neutrino in a process with a half-life of 5730 years. The isotope ¹⁴C originates in the atmosphere, due primarily to neutron bombardment of ¹⁴N caused ultimately by cosmic radiation. Because of its relatively short half-life (in geologic terms), ¹⁴C occurs at extremely low levels in fossil carbon. Over the course of 1 million years without exposure to the atmosphere, just 1 part in 10⁵⁰ will remain ¹⁴C.

The ¹³C:¹²C ratio varies slightly but measurably among natural carbon sources. Generally these differences are expressed as deviations from the ¹³C:¹²C ratio in a standard material. The international standard for carbon is Pee Dee Belemnite, a form of limestone found in South Carolina, with a ¹³C fraction of 0.0112372. For a carbon source a, the deviation of the ¹³C:¹²C ratio from that of Pee Dee Belemnite is expressed as: $\delta_a = (R_a/R_s) - 1$, where $R_a = ^{13}\text{C}:^{12}\text{C}$ ratio in the natural source, and $R_s = ^{13}\text{C}:^{12}\text{C}$ ratio in Pee Dee Belemnite, the standard. For convenience, δ_a is expressed in parts per thousand, or ‰. A negative value of δ_a shows a bias toward ¹²C over ¹³C as compared to Pee Dee Belemnite. Table A shows δ_a and ¹⁴C fraction for several natural sources of carbon.

TABLE A

13C:12C variations in natural carbon sources		
Source	- δ_a (‰)	References
Underground coal	32.5	Farquhar et al. (1989) <i>Plant Mol. Biol.</i> , 40: 503-37
Fossil fuels	26	Farquhar et al. (1989) <i>Plant Mol. Biol.</i> , 40: 503-37
Ocean DIC*	0-1.5	Goericke et al. (1994) Chapter 9 in <i>Stable Isotopes in Ecology and Environmental Science</i> , by K. Lajtha and R. H. Michener, Blackwell Publishing; Ivlev (2010) <i>Separation Sci. Technol.</i> 36: 1819-1914
Atmospheric CO ₂	6-8	Ivlev (2010) <i>Separation Sci. Technol.</i> 36: 1819-1914; Farquhar et al. (1989) <i>Plant Mol. Biol.</i> , 40: 503-37
Freshwater DIC*	6-14	Dettman et al. (1999) <i>Geochim. Cosmochim. Acta</i> 63: 1049-1057
Pee Dee Belemnite	0	Ivlev (2010) <i>Separation Sci. Technol.</i> 36: 1819-1914

*DIC = dissolved inorganic carbon

Biological processes often discriminate among carbon isotopes. The natural abundance of ¹⁴C is very small, and hence

discrimination for or against ^{14}C is difficult to measure. Biological discrimination between ^{13}C and ^{12}C , however, is well-documented. For a biological product p, we can define similar quantities to those above: $\delta_p = (R_p/R_s) - 1$, where $R_p = ^{13}\text{C}:^{12}\text{C}$ ratio in the biological product, and $R_s = ^{13}\text{C}:^{12}\text{C}$ ratio in Pee Dee Belemnite, the standard. Table B shows measured deviations in the $^{13}\text{C}:^{12}\text{C}$ ratio for some biological products.

TABLE B

$^{13}\text{C}:^{12}\text{C}$ variations in selected biological products			
Product	$-\delta_p$ (‰)	-D (‰)*	References
Plant sugar/starch from atmospheric CO_2	18-28	10-20	Ivlev (2010) <i>Separation Sci. Technol.</i> 36: 1819-1914
Cyanobacterial biomass from marine DIC	18-31	16.5-31	Goericke et al. (1994) Chapter 9 in <i>Stable Isotopes in Ecology and Environmental Science</i> , by K. Lajtha and R. H. Michener, Blackwell Publishing; Sakata et al. (1997) <i>Geochim. Cosmochim. Acta</i> , 61: 5379-89
Cyanobacterial lipid from marine DIC	39-40	37.5-40	Sakata et al. (1997) <i>Geochim. Cosmochim. Acta</i> , 61: 5379-89
Algal lipid from marine DIC	17-28	15.5-28	Goericke et al. (1994) Chapter 9 in <i>Stable Isotopes in Ecology and Environmental Science</i> , by K. Lajtha and R. H. Michener, Blackwell Publishing; Abelseon et al. (1961) <i>Proc. Natl. Acad. Sci.</i> , 47: 623-32
Algal biomass from freshwater DIC	17-36	3-30	Marty et al. (2008) <i>Limnol. Oceanogr.: Methods</i> 6: 51-63
<i>E. coli</i> lipid from plant sugar	15-27	near 0	Monson et al. (1980) <i>J. Biol. Chem.</i> , 255: 11435-41
Cyanobacterial lipid from fossil carbon	63.5-66	37.5-40	—
Cyanobacterial biomass from fossil carbon	42.5-57	16.5-31	—

*D = discrimination by a biological process in its utilization of ^{12}C vs. ^{13}C (see text)

Table B introduces a new quantity, D. This is the discrimination by a biological process in its utilization of ^{12}C vs. ^{13}C . We define D as follows: $D = (R_p/R_s) - 1$. This quantity is very similar to δ_a and δ_p , except we now compare the biological product directly to the carbon source rather than to a standard. Using D, we can combine the bias effects of a carbon source and a biological process to obtain the bias of the biological product as compared to the standard. Solving for δ_p , we obtain: $\delta_p = (D)(\delta_a) + D + \delta_a$, and, because $(D)(\delta_a)$ is generally very small compared to the other terms, $\delta_p \approx \delta_a + D$.

For a biological product having a production process with a known D, we may therefore estimate δ_p by summing δ_a and D. We assume that D operates irrespective of the carbon source. This has been done in Table B for cyanobacterial lipid and biomass produced from fossil carbon. As shown in the Table A and Table B, above, cyanobacterial products made from fossil carbon (in the form of, for example, flue gas or other emissions) will have a higher δ_p than those of comparable biological products made from other sources, distinguishing them on the basis of composition of matter from these other biological products. In addition, any product derived solely from fossil carbon will have a negligible frac-

tion of ^{14}C , while products made from above-ground carbon will have a ^{14}C fraction of approximately 10^{-12} .

Accordingly, in certain aspects, the invention provides various carbon-based products of interest characterized as $-\delta_p$ (‰) of about 63.5 to about 66 and $-D$ (‰) of about 37.5 to about 40.

Antibodies

In another aspect, the present invention provides isolated antibodies, including fragments and derivatives thereof that bind specifically to the isolated polypeptides and polypeptide fragments of the present invention or to one or more of the polypeptides encoded by the isolated nucleic acids of the present invention. The antibodies of the present invention may be specific for linear epitopes, discontinuous epitopes or conformational epitopes of such polypeptides or polypeptide fragments, either as present on the polypeptide in its native conformation or, in some cases, as present on the polypeptides as denatured, as, e.g., by solubilization in SDS. Among the useful antibody fragments provided by the instant invention are Fab, Fab', Fv, F(ab')₂, and single chain Fv fragments.

By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

As is well known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to unrelated polypeptides by at least two-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a polypeptide or polypeptide fragment of the present invention will be at least about 1×10^{-6} M, typically at least about 5×10^{-7} M, usefully at least about 1×10^{-7} M, with affinities and avidities of 1×10^{-8} M, 5×10^{-9} M, 1×10^{-10} M and even stronger proving especially useful.

The isolated antibodies of the present invention may be naturally-occurring forms, such as IgG, IgM, IgD, IgE, and IgA, from any mammalian species. For example, antibodies are usefully obtained from species including rodents—typically mouse, but also rat, guinea pig, and hamster-lagomorphs, typically rabbits, and also larger mammals, such as sheep, goats, cows, and horses. The animal is typically affirmatively immunized, according to standard immunization protocols, with the polypeptide or polypeptide fragment of the present invention.

Virtually all fragments of 8 or more contiguous amino acids of the polypeptides of the present invention may be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker. Immunogenicity may also be conferred by fusion of the polypeptide and polypeptide fragments of the present invention to other moieties. For example, peptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine

development. See, e.g., Tam et al., *Proc. Natl. Acad. Sci. USA* 85:5409-5413 (1988); Posnett et al., *J. Biol. Chem.* 263, 1719-1725 (1988).

Protocols for immunization are well-established in the art. Such protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant. Antibodies of the present invention may be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immuno-histochemical detection of the proteins of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention. Following immunization, the antibodies of the present invention may be produced using any art-accepted technique. Host cells for recombinant antibody production—either whole antibodies, antibody fragments, or antibody derivatives—can be prokaryotic or eukaryotic. Prokaryotic hosts are particularly useful for producing phage displayed antibodies, as is well known in the art. Eukaryotic cells, including mammalian, insect, plant and fungal cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention. Antibodies of the present invention can also be prepared by cell free translation.

The isolated antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the polypeptides and polypeptide fragments of the present invention. The choice of label depends, in part, upon the desired use. In some cases, the antibodies of the present invention may usefully be labeled with an enzyme. Alternatively, the antibodies may be labeled with colloidal gold or with a fluorophore. For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention may usefully be labeled with biotin. When the antibodies of the present invention are used, e.g., for Western blotting applications, they may usefully be labeled with radioisotopes, such as ³³P, ³²P, ³⁵S, ³H and ¹²⁵I. As would be understood, use of the labels described above is not restricted to any particular application.

The following examples are for illustrative purposes and are not intended to limit the scope of the present invention.

Example 1

A Pathway for the Enzymatic Synthesis of n-Alkanes

An enzymatic process for the production of n-alkanes in, e.g., cyanobacteria is shown in FIG. 1A based on the sequential activity of (1) an AAR enzyme, e.g., tll1312, an acyl-ACP reductase; and (2) an ADM enzyme, e.g., tll1313, a putative alkanal decarboxylative monooxygenase, that uses reduced ferredoxin as electron donor. The AAR activity is distinct from the relatively well characterized acyl-CoA reductase activity exhibited by proteins such as Acr1 from *Acinetobacter calcoaceticus* (Reiser S and Somerville C (1997) *J. Bacteriol.* 179:2969-2975). A membranous ADM activity has previously been identified in insect microsomal preparations (Reed JR et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:10000-10004; Reed JR et al. (1995) *Musca domestica. Biochemistry* 34:16221-16227).

FIGS. 1B and 1C summarize the names and activities of the enzymes involved in the biosynthesis of n-alkanals. FIG. 1B depicts the relatively well characterized acyl-CoA reductase activity (EC 1.2.1.50) exhibited by proteins such as Acr1 from *Acinetobacter calcoaceticus*. In FIG. 1C, the two well-known

ACP-related reductases that are involved in fatty acid biosynthesis, β -ketoacyl-ACP reductase (EC 1.1.1.100) and enoyl-ACP reductase (EC 1.3.1.9, 1.3.1.10), are contrasted with the acyl-ACP reductase (AAR) (no EC number yet assigned) believed to be involved in the biosynthetic pathway for n-alkanes in cyanobacteria. The key difference between AAR and acyl-CoA reductase (EC 1.2.1.50) is that ACP is the acyl carrier rather than coenzyme A. Supporting this distinction, it has been shown that acyl-CoA reductase Acr1 from *Acinetobacter calcoaceticus* can only generate alkanals from acyl-CoA and not acyl-ACP (Resier S and Somerville C (1997) *J. Bacteriol.* 179: 2969-2975).

ADM also lacks a presently assigned EC number. An alkanal monooxygenase (EC 1.14.14.3), often referred to as luciferase, is known to catalyze the conversion of n-alkanal to n-alkanoic acid. This activity is distinct from the ADM activity (n-alkanal to (n-1)-alkane) proposed herein, although both use n-alkanal and molecular oxygen as substrates.

Cyanobacterial AAR and ADM Homologs for Production of n-Alkanes.

In this example, homologs of cyanobacterial AAR and ADM genes (e.g., homologs of *Synechococcus elongatus* PCC 7942 SYNPC7942_1594 and/or SYNPC7942_1593 protein, respectively) are identified using a BLAST search. These proteins can be expressed in a variety of organisms (bacteria, yeast, plant, etc.) for the purpose of generating and isolating n-alkanes and other desired carbon-based products of interest from the organisms. A search of the non-redundant BLAST protein database revealed counterparts for each protein in other cyanobacteria.

To determine the degree of similarity among homologs of the *Synechococcus elongatus* PCC 7942 SYNPC7942_1594 protein, the 341-amino acid protein sequence was queried using BLAST (<http://blast.ncbi.nlm.nih.gov/>) against the "nr" non-redundant protein database. Homologs were taken as matching proteins whose alignments (i) covered >90% length of SYNPC7942_1594, (ii) covered >90% of the length of the matching protein, and (iii) had >50% identity with SYNPC7942_1594 (Table 1).

TABLE 1

Protein homologs of SYNPC7942_1594 (AAR)			
Organism	SEQ ID NO: Homolog accession #	BLAST Score, E-value	
<i>Synechococcus elongatus</i> PCC 7942	6 (SYNPC7942_1594)	n/a	
<i>Synechococcus elongatus</i> PCC 7942 [cyanobacteria] taxid 1140	23 YP_400611.1	706, 0.0	
<i>Synechococcus elongatus</i> PCC 6301 [cyanobacteria] taxid 269084	24 YP_170761.1	706, 0.0	
<i>Anabaena variabilis</i> ATCC 29413 [cyanobacteria] taxid 240292	25 YP_323044.1	538, 4e-151	
<i>Nostoc</i> sp. PCC 7120 [cyanobacteria] taxid 103690	26 NP_489324.1	535, 3e-150	
<i>Nostoc azollae</i> 0708 [cyanobacteria] taxid 551115	27 ZP_03763674.1	533, 1e-149	
<i>Cyanospora</i> sp. PCC 7425 [cyanobacteria] taxid 395961	28 YP_002481152.1	526, 9e-148	
<i>Nodularia spumigena</i> CCY 9414 [cyanobacteria] taxid 313624	29 ZP_01628095.1	521, 3e-146	

33

TABLE 1-continued

Protein homologs of SYNPPCC7942_1594 (AAR)			
Organism	SEQ ID NO: Homolog accession #	BLAST Score, E-value	
<i>Lyngbya</i> sp. PCC 8106 [cyanobacteria] taxid 313612	30 ZP_01619574.1	520, 6e-146	
<i>Nostoc punctiforme</i> PCC 73102 [cyanobacteria] taxid 63737	31 YP_001865324.1	520, 7e-146	
<i>Trichodesmium erythraeum</i> IMS101 [cyanobacteria] taxid 203124	32 YP_721978.1	517, 6e-145	
<i>Thermosynechococcus elongatus</i> BP-1 [cyanobacteria] taxid 197221	2 NP_682102.1	516, 2e-144	
<i>Acaryochloris marina</i> MBIC11017 [cyanobacteria] taxid 329726	33 YP_001518341.1	512, 2e-143	
<i>Cyanothece</i> sp. PCC 8802 [cyanobacteria] taxid 395962	34 ZP_03142196.1	510, 8e-143	
<i>Cyanothece</i> sp. PCC 8801 [cyanobacteria] taxid 41431	35 YP_002371106.1	510, 8e-143	
<i>Microcoleus chthonoplastes</i> PCC 7420 [cyanobacteria] taxid 118168	36 YP_002619867.1	509, 2e-142	
<i>Arthrospira maxima</i> CS-328 [cyanobacteria] taxid 513049	37 ZP_03273554.1	507, 7e-142	
<i>Synechocystis</i> sp. PCC 6803 [cyanobacteria] taxid 1148	38 NP_442146.1	504, 5e-141	
<i>Cyanothece</i> sp. CCY 0110 [cyanobacteria] taxid 391612	39 ZP_01728620.1	501, 4e-140	
<i>Synechococcus</i> sp. PCC 7335 [cyanobacteria] taxid 91464	40 YP_002711557.1	500, 1e-139	
<i>Cyanothece</i> sp. ATCC 51142 [cyanobacteria] taxid 43989	41 YP_001802846.1	489, 2e-136	
<i>Gloeobacter violaceus</i> PCC 7421 [cyanobacteria] taxid 251221	42 NP_926091.1	487, 7e-136	
<i>Microcystis aeruginosa</i> NIES-843 [cyanobacteria] taxid 449447	43 YP_001660322.1	486, 1e-135	
<i>Crocospaera watsonii</i> WH 8501 [cyanobacteria] taxid 165597	44 ZP_00516920.1	486, 1e-135	
<i>Microcystis aeruginosa</i> PCC 7806 [cyanobacteria] taxid 267872	45 emb CAO90781.1	484, 8e-135	
<i>Synechococcus</i> sp. WH 5701 [cyanobacteria] taxid 69042	46 ZP_01085337.1	471, 4e-131	
<i>Synechococcus</i> sp. RCC307 [cyanobacteria] taxid 316278	47 YP_001227841.1	464, 8e-129	
uncultured marine type-A <i>Synechococcus</i> GOM 3O6 [cyanobacteria] taxid 364150	48 gb ABD96327.1	462, 2e-128	
<i>Synechococcus</i> sp. WH 8102 [cyanobacteria] taxid 84588	49 NP_897828.1	462, 2e-128	

34

TABLE 1-continued

Protein homologs of SYNPPCC7942_1594 (AAR)			
Organism	SEQ ID NO: Homolog accession #	BLAST Score, E-value	
<i>Synechococcus</i> sp. WH 7803 [cyanobacteria] taxid 32051	50 YP_001224378.1	459, 2e-127	
uncultured marine type-A <i>Synechococcus</i> GOM 5D20 [cyanobacteria] taxid 364154	51 gb ABD96480.1	458, 3e-127	
<i>Synechococcus</i> sp. WH 7805 [cyanobacteria] taxid 59931	52 ZP_01123215.1	457, 5e-127	
uncultured marine type-A <i>Synechococcus</i> 5B2 [cyanobacteria] taxid 359140	53 gb ABB92249.1	457, 8e-127	
<i>Synechococcus</i> sp. RS9917 [cyanobacteria] taxid 221360	54 ZP_01079773.1	456, 2e-126	
<i>Synechococcus</i> sp. CC9902 [cyanobacteria] taxid 316279	55 YP_377636.1	454, 6e-126	
<i>Prochlorococcus marinus</i> subsp. <i>marinus</i> str. CCMP1375 [cyanobacteria] taxid 167539	56 NP_874926.1	453, 9e-126	
<i>Prochlorococcus marinus</i> str. MIT 9313 [cyanobacteria] taxid 74547	57 NP_895058.1	453, 1e-125	
uncultured marine type-A <i>Synechococcus</i> GOM 3M9 [cyanobacteria] taxid 364149	58 gb ABD96274.1	452, 2e-125	
uncultured marine type-A <i>Synechococcus</i> GOM 4P21 [cyanobacteria] taxid 364153	59 gb ABD96442.1	452, 2e-125	
<i>Synechococcus</i> sp. BL107 [cyanobacteria] taxid 313625	60 ZP_01469469.1	452, 2e-125	
<i>Cyanobium</i> sp. PCC 7001 [cyanobacteria] taxid 180281	61 YP_002597253.1	451, 4e-125	
<i>Prochlorococcus marinus</i> str. NATL1A [cyanobacteria] taxid 167555	62 YP_001014416.1	449, 2e-124	
<i>Prochlorococcus marinus</i> str. MIT 9515 [cyanobacteria] taxid 167542	63 YP_001010913.1	447, 6e-124	
<i>Synechococcus</i> sp. CC9605 [cyanobacteria] taxid 110662	64 YP_381056.1	447, 8e-124	
<i>Prochlorococcus marinus</i> str. MIT 9211 [cyanobacteria] taxid 93059	65 YP_001550421.1	446, 2e-123	
<i>Prochlorococcus marinus</i> subsp. <i>pastoris</i> str. CCMP1986 [cyanobacteria] taxid 59919	66 NP_892651.1	446, 2e-123	
<i>Prochlorococcus marinus</i> str. MIT 9301 [cyanobacteria] taxid 167546	67 YP_001090783.1	445, 3e-123	
<i>Synechococcus</i> sp. RS9916 [cyanobacteria] taxid 221359	68 ZP_01472595.1	445, 3e-123	

35

TABLE 1-continued

Protein homologs of SYNPPCC7942_1594 (AAR)			
Organism	SEQ ID NO:	Homolog accession #	BLAST Score, E-value
<i>Prochlorococcus marinus</i> str. NATL2A [cyanobacteria] taxid 59920	69	YP_293055.1	445, 4e-123
<i>Prochlorococcus marinus</i> str. MIT 9202 [cyanobacteria] taxid 93058	70	YP_002673377.1	444, 7e-123
<i>Synechococcus</i> sp. CC9311 [cyanobacteria] taxid 64471	71	YP_731192.1	443, 1e-122
<i>Prochlorococcus marinus</i> str. MIT 9215 [cyanobacteria] taxid 93060	72	YP_001483815.1	442, 2e-122
<i>Prochlorococcus marinus</i> str. AS9601 [cyanobacteria] taxid 146891	73	YP_001008982.1	442, 3e-122
<i>Synechococcus</i> sp. JA-3-3Ab [cyanobacteria] taxid 321327	74	YP_473896.1	441, 5e-122
<i>Synechococcus</i> sp. JA-2-3B'a(2-13) [cyanobacteria] taxid 321332	75	YP_478638.1	440, 8e-122
<i>Prochlorococcus marinus</i> str. MIT 9312 [cyanobacteria] taxid 74546	76	YP_397030.1	436, 1e-120

To determine the degree of similarity among homologs of the *Synechococcus elongatus* PCC 7942 SYNPPCC7942_1593 protein, the 231 amino acid protein sequence was queried using BLAST (<http://blast.ncbi.nlm.nih.gov/>) against the "nr" non-redundant protein database. Homologs were taken as matching proteins whose alignments (i) covered >90% length of SYNPPCC7942_1593, (ii) covered >90% of the length of the matching protein, (iii) and had >50% identity with SYNPPCC7942_1593 (Table 2).

TABLE 2

Protein homologs of SYNPPCC7942_1593 (ADM)			
Organism	SEQ ID NO:	Homolog accession #	BLAST Score, E-value
<i>Synechococcus elongatus</i> PCC 7942 [cyanobacteria]	8	(SYNPPCC7942_1593)	n/a
<i>Synechococcus elongatus</i> PCC 7942 [cyanobacteria] taxid 1140	77	YP_400610.1	475, 1e-132
<i>Synechococcus elongatus</i> PCC 6301 [cyanobacteria] taxid 269084	78	YP_170760.1	475, 2e-132
<i>Arthrospira maxima</i> CS-328 [cyanobacteria] taxid 513049	79	ZP_03273549.1	378, 3e-103
<i>Microcoleus chthonoplastes</i> PCC 7420 [cyanobacteria] taxid 118168	80	YP_002619869.1	376, 1e-102
<i>Lyngbya</i> sp. PCC 8106 [cyanobacteria] taxid 313612	81	ZP_01619575.1	374, 5e-102
<i>Nodularia spumigena</i> CCY 9414 [cyanobacteria] taxid 313624	82	ZP_01628096.1	369, 1e-100

36

TABLE 2-continued

Protein homologs of SYNPPCC7942_1593 (ADM)			
Organism	SEQ ID NO:	Homolog accession #	BLAST Score, E-value
<i>Microcystis aeruginosa</i> NIES-843 [cyanobacteria] taxid 449447	83	YP_001660323.1	367, 5e-100
<i>Microcystis aeruginosa</i> PCC 7806 [cyanobacteria] taxid 267872	84	emb1CAO90780.1	364, 3e-99
<i>Nostoc</i> sp. PCC 7120 [cyanobacteria] taxid 103690	85	NP_489323.1	363, 1e-98
<i>Anabaena variabilis</i> ATCC 29413 [cyanobacteria] taxid 240292	86	YP_323043.1	362, 2e-98
<i>Crocospaera watsonii</i> WH 8501 [cyanobacteria] taxid 165597	87	ZP_00514700.1	359, 1e-97
<i>Trichodesmium erythraeum</i> IMS101 [cyanobacteria] taxid 203124	88	YP_721979.1	358, 2e-97
<i>Synechococcus</i> sp. PCC 7335 [cyanobacteria] taxid 91464	89	YP_002711558.1	357, 6e-97
' <i>Nostoc azollae</i> ' 0708 [cyanobacteria] taxid 551115	90	ZP_03763673.1	355, 3e-96
<i>Synechocystis</i> sp. PCC 6803 [cyanobacteria] taxid 1148	91	NP_442147.1	353, 5e-96
<i>Cyanothece</i> sp. ATCC 51142 [cyanobacteria] taxid 43989	92	YP_001802195.1	352, 2e-95
<i>Cyanothece</i> sp. CCY 0110 [cyanobacteria] taxid 391612	93	ZP_01728578.1	352, 2e-95
<i>Cyanothece</i> sp. PCC 7425 [cyanobacteria] taxid 395961	94	YP_002481151.1	350, 7e-95
<i>Nostoc punctiforme</i> PCC 73102 [cyanobacteria] taxid 63737	95	YP_001865325.1	349, 1e-94
<i>Acaryochloris marina</i> MBIC11017 [cyanobacteria] taxid 329726	96	YP_001518340.1	344, 4e-93
<i>Cyanothece</i> sp. PCC 8802 [cyanobacteria] taxid 395962	97	ZP_03142957.1	342, 1e-92
<i>Cyanothece</i> sp. PCC 8801 [cyanobacteria] taxid 41431	98	YP_002370707.1	342, 1e-92
<i>Thermosynechococcus elongatus</i> BP-1 [cyanobacteria] taxid 197221	4	NP_682103.1	332, 2e-89
<i>Synechococcus</i> sp. JA-2-3B'a(2-13) [cyanobacteria] taxid 321332	50		
<i>Synechococcus</i> sp. RCC307 [cyanobacteria] taxid 316278	55		
<i>Synechococcus</i> sp. WH 7803 [cyanobacteria] taxid 32051	60		
<i>Synechococcus</i> sp. WH 8102 [cyanobacteria] taxid 84588	100	YP_001227842.1	319, 1e-85
<i>Synechococcus</i> sp. WH 7805 [cyanobacteria] taxid 59931	101	YP_001224377.1	313, 8e-84
uncultured marine type-A	102	NP_897829.1	311, 3e-83
	103	ZP_01123214.1	310, 6e-83

TABLE 2-continued

Protein homologs of SYNPC927942_1593 (ADM)			
Organism	SEQ ID NO:	Homolog accession #	BLAST Score, E-value
<i>Synechococcus</i> GOM 3O12 [cyanobacteria] taxid 364151	104	gb ABD96376.1	309, 1e-82
<i>Synechococcus</i> sp. JA-3-3Ab [cyanobacteria] taxid 321327 uncultured marine type-A	105	YP_473897.1	309, 1e-82
<i>Synechococcus</i> GOM 3O6 [cyanobacteria] taxid 364150 uncultured marine type-A	106	gb ABD96328.1	309, 1e-82
<i>Synechococcus</i> GOM 3M9 [cyanobacteria] taxid 364149	107	gb ABD96275.1	308, 2e-82
<i>Synechococcus</i> sp. CC9311 [cyanobacteria] taxid 64471 uncultured marine type-A	108	YP_731193.1	306, 7e-82
<i>Synechococcus</i> 5B2 [cyanobacteria] taxid 359140	109	gb ABB92250.1	306, 9e-82
<i>Synechococcus</i> sp. WH 5701 [cyanobacteria] taxid 69042	110	ZP_01085338.1	305, 3e-81
<i>Gloeobacter violaceus</i> PCC 7421 [cyanobacteria] taxid 251221	111	NP_926092.1	303, 8e-81
<i>Synechococcus</i> sp. RS9916 [cyanobacteria] taxid 221359	112	ZP_01472594.1	303, 9e-81
<i>Synechococcus</i> sp. RS9917 [cyanobacteria] taxid 221360	113	ZP_01079772.1	300, 6e-80
<i>Synechococcus</i> sp. CC9605 [cyanobacteria] taxid 110662	114	YP_381055.1	300, 7e-80
<i>Prochlorococcus marinus</i> str. MIT 9303 [cyanobacteria] taxid 59922	115	YP_001016795.1	294, 4e-78
<i>Cyanobium</i> sp. PCC 7001 [cyanobacteria] taxid 180281	116	YP_002597252.1	294, 6e-78
<i>Prochlorococcus marinus</i> str. MIT 9313 [cyanobacteria] taxid 74547	117	NP_895059.1	291, 3e-77
<i>Synechococcus</i> sp. CC9902 [cyanobacteria] taxid 316279	118	YP_377637.1	289, 1e-76
<i>Prochlorococcus marinus</i> str. MIT 9301 [cyanobacteria] taxid 167546	119	YP_001090782.1	287, 5e-76
<i>Synechococcus</i> sp. BL107 [cyanobacteria] taxid 313625	120	ZP_01469468.1	287, 6e-76
<i>Prochlorococcus marinus</i> str. AS9601 [cyanobacteria] taxid 146891	121	YP_001008981.1	286, 2e-75
<i>Prochlorococcus marinus</i> str. MIT 9312 [cyanobacteria] taxid 74546	12	YP_397029.1	282, 1e-74
<i>Prochlorococcus marinus</i> subsp. <i>pastoris</i> str. CCMP1986 [cyanobacteria] taxid 59919	122	NP_892650.1	280, 9e-74

TABLE 2-continued

Protein homologs of SYNPC927942_1593 (ADM)			
Organism	SEQ ID NO:	Homolog accession #	BLAST Score, E-value
<i>Prochlorococcus marinus</i> str. MIT 9211 [cyanobacteria] taxid 93059	123	YP_001550420.1	279, 2e-73
<i>Prochlorococcus marinus</i> str. NATL2A [cyanobacteria] taxid 59920	124	YP_293054.1	276, 9e-73
<i>Prochlorococcus marinus</i> str. NATL1A [cyanobacteria] taxid 167555	125	YP_001014415.1	276, 9e-73
<i>Prochlorococcus marinus</i> subsp. <i>marinus</i> str. CCMP1375 [cyanobacteria] taxid 167539	126	NP_874925.1	276, 1e-72
<i>Prochlorococcus marinus</i> str. MIT 9515 [cyanobacteria] taxid 167542	127	YP_001010912.1	273, 6e-72
<i>Prochlorococcus marinus</i> str. MIT 9215 [cyanobacteria] taxid 93060	128	YP_001483814.1	273, 9e-72

The amino acid sequences referred to in the Table, as those sequences appeared in the NCBI database on Jul. 9, 2009, by accession number are incorporated by reference herein.

An AAR enzyme from Table 1, and/or an ADM enzyme from Table 2, or both can be expressed in a host cell of interest, wherein the host may be a heterologous host or the native host, i.e., the species from which the genes were originally derived. In one embodiment, the invention provides a method of imparting n-alkane synthesis capability in a heterologous organism, lacking native homologs of AAR and/or ADM, by engineering the organism to express a gene encoding one of the enzymes listed in Table 1 or Table 2. Also provided are methods of modulating n-alkane synthesis in an organism which already expresses one or both of the AAR and ADM enzymes by increasing the expression of the native enzymes, or by augmenting native gene expression by the recombinant expression of heterologous AAR and/or ADM enzymes. In addition, the invention provides methods of modulating the degree of alkane synthesis by varying certain parameters, including the identity and/or compatibility of electron donors, culture conditions, promoters for expressing AAR and/or ADM enzymes, and the like.

If the host lacks a suitable electron donor or lacks sufficient levels of a suitable electron donor to achieve production of the desired amount of n-alkane, such electron donor may also be introduced recombinantly. Guidelines for optimizing electron donors for the reaction catalyzed by the recombinant ADM proteins described herein may be summarized as follows:

1. In cyanobacteria, electrons are shuttled from photosystem I to ferredoxin and from ferredoxin to the ADM enzyme.
2. In bacteria that lack photosystem I, electrons can be shuttled from NADPH to ferredoxin via the action of ferredoxin-NADP+ reductase (EC 1.18.1.2) and from ferredoxin to the ADM enzyme.
3. In bacteria that lack photosystem I, electrons can be shuttled from NADPH to flavodoxin via the action of

ferredoxin-NADP⁺ reductase (EC 1.18.1.2) and from flavodoxin to the ADM enzyme.

4. In bacteria that lack photosystem I, electrons can be shuttled from NADH to ferredoxin via the action of *Trichomonas vaginalis* NADH dehydrogenase and from ferredoxin to the ADM enzyme.
5. In all bacteria, electrons can be shuttled from pyruvate to ferredoxin by the action of pyruvate:ferredoxin oxidoreductase (EC 1.2.7.1), and from ferredoxin to the ADM enzyme.

In addition to the *in vivo* production of n-alkanes discussed above, AAR and ADM proteins encoded by the genes listed in Tables 1 and 2 can be purified. When incubated *in vitro* with an appropriate electron donor (e.g., a ferredoxin, as discussed above), the proteins will catalyze the enzymatic synthesis of n-alkanes *in vitro* from appropriate starting materials (e.g., an acyl-ACP or n-alkanal).

In addition to the pathways for n-alkane synthesis described above, the invention also provides an alternative pathway, namely, acyl-CoA → n-alkanal → (n-1)-alkane, via the successive activities of acyl-CoA reductase (ACR) and ADM. Normally, acyl-CoA is the first intermediate in metabolic pathways of fatty acid oxidation; thus, upon import into the cell, exogenously added free fatty acids are converted to acyl-CoAs by acyl-CoA synthetase (FIG. 1B). Acyl-CoA can also be derived purely biosynthetically as follows: acyl-ACP → free fatty acid → acyl-CoA, via the activities of cytoplasmic acyl-ACP thioesterase (EC 3.1.2.14; an example is leader-signal-less *E. coli* TesA) and the endogenous and/or heterologous acyl-CoA synthetase. Thus, in one embodiment, the invention provides a method for the biosynthesis of n-alkanes via the pathway: acyl-ACP → intracellular free fatty acid → acyl-CoA → n-alkanal → (n-1)-alkane (FIG. 1D), catalyzed by the successive activities of acyl-ACP thioesterase, acyl-CoA synthetase, acyl-CoA reductase, and ADM. For example, the acyl-CoA reductase Acr1 from *Acinetobacter calcoaceticus* and the ADM from *Synechococcus* sp. PCC7942 (SYNPCC7942_1593) can be used to transform *E. coli*, which is cultured in the presence of exogenous free fatty acids. The free fatty acids are taken up by the cells as acyl-CoA, which are then converted to n-alkanal by Acr1, and thence to (n-1)-alkane by ADM.

Example 2

Production of n-Alkanes, n-Alkenes, and Fatty Alcohols in *Escherichia coli* K-12 Through Heterologous Expression of *Synechococcus elongatus* PCC7942 SYNPCC7942_1593 (adm) and SYNPCC7942_594 (aar)

The natural SYNPCC7942_1593-SYNPCC7942_1594 operonic sequence was PCR-amplified from the genomic DNA of *Synechococcus elongatus* PCC7942 and cloned into the pAQ1 homologous recombination vector pJB5 via NdeI and EcoRI. The resulting plasmid was denoted pJB823. This construct placed the SYNPCC7942_1593-SYNPCC7942_1594 operon under the transcriptional control of the constitutive aphII promoter. The sequence of pJB823 is provided as SEQ ID NO: 15. The intracellular hydrocarbon products of *E. coli* K-12 EPI400™ (Epicentre) harboring pJB823, JCC1076, were compared to those of EPI400™ harboring pJB5, the control strain JCC9a, by gas chromatography-mass spectrometry (GC-MS). Clonal cultures of JCC9a and JCC1076 were grown overnight at 37° C. in Luria Broth (LB) containing 2% glucose, 100 µg/ml carbenicillin, 50 µg/ml spectinomycin, 50 µg/ml streptomycin, and 1× CopyCutter

Induction Solution (Epicentre). For each strain, 15 ml of saturated culture was collected by centrifugation. Cell pellets were washed thoroughly by three cycles of resuspension in Milli-Q water and microcentrifugation, and then dewatered as much as possible by three cycles of microcentrifugation and aspiration. Cell pellets were then extracted by vortexing for five minutes in 0.8 ml acetone containing 100 µg/ml butylated hydroxytoluene (BHT; a general antioxidant) and 100 µg/ml ethyl arachidate (EA; an internal reporter of extraction efficiency). Cell debris was pelleted by centrifugation, and 700 µl extractant was pipetted into a GC vial. These JCC9a and JCC1076 acetone samples, along with authentic standards, were then analyzed by GC-MS.

The gas chromatograph was an Agilent 7890A GC equipped with a 5975C electron-impact mass spectrometer. Liquid samples (1.0 µl) were injected into the GC with a 7683 automatic liquid sampler equipped with a 10 µl syringe. The GC inlet temperature was 290° C. and split-less injection was used. The capillary column was an Agilent HP-5MS (30 m×0.25 mm×0.25 µm). The carrier gas was helium at a flow rate of 1.0 ml/min. The GC oven temperature program was 50° C., hold 1 min/10° C. per min to 290° C./hold 9 min. The GC-MS interface temperature was 290° C. The MS source temperature was 230° C., and the quadrupole temperature was 150° C. The mass range was 25-600 amu. MS fragmentation spectra were matched against the NIST MS database, 2008 version.

Peaks present in the total-ion GC-MS chromatograms were chemically assigned in one of two ways. In the first, assignment was done by ensuring that both the retention time and the fragmentation mass spectrum corresponded to the retention time and fragmentation mass spectrum, respectively, of an authentic standard—this is referred to as “Method 1”, and is essentially unambiguous. In the absence of authentic standards, only a tentative chemical assignment can be reached; this was done by collectively integrating the following data for the peak in question: (i) the structure of the fragmentation spectrum, especially with regard to the weight of the molecular ion, and to the degree to which it resembled a hydrocarbon-characteristic “envelope” mass spectrum, (ii) the retention time, especially with regard to its qualitative compatibility with the assigned compound, e.g., cis-unsaturated n-alkenes elute slightly before their saturated n-alkane counterparts, and (iii) the likelihood that the assigned compound is chemically compatible with the operation of the AAR-ADM and related pathways in the host organism in question, e.g., fatty aldehydes generated by AAR are expected to be converted to the corresponding fatty alcohols by host dehydrogenases in *E. coli* if they are not acted upon sufficiently quickly by ADM. This second approach to peak assignment is referred to as “Method 2”. In the total-ion GC-MS chromatogram in FIG. 2, as well as in all such chromatograms in subsequent figures, peaks chemically assigned by Method 1 are labeled in regular font, whereas those assigned by Method 2 are labeled in italic font.

Total ion chromatograms (TICs) of JCC9a and JCC1076 acetone cell pellet extractants are shown in FIG. 2. The TICs of C₈-C₂₀ n-alkane authentic standards (Sigma 04070), as well as 1-tetradecanol (Sigma 185388) plus 1-hexadecanol (Sigma 258741) plus 1-octadecanol (Sigma 258768), are also shown. Hydrocarbons identified in JCC1076, but not in control strain JCC9a, are detailed in Table 3. These hydrocarbons are n-pentadecane (1), 1-tetradecanol (1), n-heptadecene (2), n-heptadecane (1), and 1-hexadecanol (1), where the number in parentheses indicates the GC-MS peak assignment method. MS fragmentation spectra of the Method 1 peaks are shown in FIG. 3, plotted against their respective library hits.

41

TABLE 3

Hydrocarbons detected by GC-MS in acetone cell pellet extractants of JCC1076 but not JCC9a, in increasing order of retention time.				
Compound	JCC9a	JCC1076	GC-MS Peak Assignment	Candidate isomer
n-pentadecane	-	+	Method 1	
1-tetradecanol	-	+	Method 1	
n-heptadecene	-	+	Method 2 (envelope-type MS with molecular ion mass 238)	cis-7-heptadecene
n-pentadecane	-	+	Method 1	
1-hexadecanol	-	+	Method 1	

"-" not detected;

"+" detected.

The formation of these five products is consistent with both the expected incomplete operation, i.e., acyl-ACP→fatty aldehyde→fatty alcohol, and expected complete operation, i.e., acyl-ACP→fatty aldehyde→alkane/alkene, of the AAR-ADM pathway in *E. coli*, whose major straight-chain acyl-ACPs include 12:0, 14:0, 16:0, 18:0, 16:1Δ9cis, and 18:1Δ11cis acyl groups (Heipieper H J (2005); *Appl Environ Microbiol* 71:3388). Assuming that n-heptadecene (2) is derived 18:1Δ11cis-ACP, it would correspond to cis-7-heptadecene. Indeed, an n-heptadecene isomer was identified as the highest-confidence MS fragmentation library hit at that retention time, with the expected molecular ion of molecular weight 238; also, as expected, it elutes slightly before n-heptadecane.

Example 3

Production of n-Alkanes, n-Alkenes, and Fatty Alcohols in *Escherichia coli* B Through Heterologous Expression of *Synechococcus elongatus* PCC7942 SYN-PCC7942_1593 (adm) and SYN-PCC7942_1594 (aar)

The natural SYN-PCC7942_1593-SYN-PCC7942_594 operonic sequence was excised from pJB823 using NdeI and EcoRI, and cloned into the commercial expression vector pCDFDuet™-1 (Novagen) cut with via NdeI and MfeI. The resulting plasmid was denoted pJB855 (SEQ ID NO: 16). This construct placed the SYN-PCC7942_1593-SYN-PCC7942_1594 operon under the transcriptional control of the inducible T7lacO promoter.

The intracellular hydrocarbon products of *E. coli* BL21 (DE3) (Novagen) harboring pJB855, JCC1113, were compared to those of *E. coli* BL21(DE3) harboring pCDFDuet™-1, the control strain JCC114, by gas chromatography-mass spectrometry (GC-MS). Starter clonal cultures of JCC1114 and JCC1113 were grown overnight at 37° C. in M9 minimal medium supplemented with 6 mg/l FeSO₄·7H₂O, 50 μg/ml spectinomycin, and 2% glucose as carbon source; this medium is referred to M9fs. Each starter culture was used to inoculate a 32 ml culture of M9fs at an initial OD₆₀₀ of 0.1. Inoculated cultures were grown at 37° C. at 300 rpm until an OD₆₀₀ of 0.4 has been reached, at which point IPTG was added to a final concentration of 1 mM. After addition of inducer, cultures were grown under the same conditions for an additional 17 hours. For each strain, 12 ml of saturated culture was then collected by centrifugation. Cell pellets were washed thoroughly by 3 cycles of resuspension in Milli-Q water and microcentrifugation, and then dewetted as much as possible by 3 cycles of microcentrifugation and aspiration.

42

Cell pellets were then extracted by vortexing for 5 minutes in 0.7 ml acetone containing 20 μg/ml BHT and 20 μg/ml EA. Cell debris was pelleted by centrifugation, and 600 μl supernatant was pipetted into a GC vial. These JCC1114 and JCC1113 samples, along with authentic standards, were then analyzed by GC-MS as described in Example 2. The TICs of JCC1114 and JCC1113 acetone cell pellet extractants are shown in FIG. 4; n-alkane and 1-alkanol standards are as in Example 2. Hydrocarbons identified in JCC1113, but not in control strain JCC1114, are detailed in Table 4.

TABLE 4

Hydrocarbons detected by GC-MS in acetone cell pellet extractants of JCC1113 but not JCC1114 in increasing order of retention time.				
Compound	JCC1114	JCC1113	GC-MS Peak Assignment	Candidate isomer
n-tridecane	-	+	Method 1	
n-tetradecane	-	+	Method 1	
n-pentadecene	-	+	Method 2 (envelope-type MS with molecular ion mass 210)	cis-7-pentadecene
1-dodecanol	-	+	Method 2	
n-pentadecane	-	+	Method 1	
n-hexadecene	-	+	Method 2 (envelope-type MS with molecular ion mass 224)	cis-8-hexadecene
n-hexadecane	-	+	Method 1	
1-tetradecanol	-	+	Method 1	
n-heptadecene	-	+	Method 2 (envelope-type MS with molecular ion mass 238)	cis-7-heptadecene
n-heptadecane	-	+	Method 1	
1-pentadecanol	-	+	Method 2	
1-hexadecanol	-	+	Method 2	cis-9-hexadecan-1-ol
1-hexadecanol	-	+	Method 1	
1-octadecanol	-	+	Method 2 (envelope-type MS with molecular ion mass 250)	cis-11-octadecan-1-ol

"-" not detected;

"+" detected.

These hydrocarbons are n-tridecane (1), n-tetradecane (1), n-pentadecene (2), 1-dodecanol (2), n-pentadecane (1), n-hexadecene (2), n-hexadecane (1), 1-tetradecanol (1), n-heptadecene (2), n-heptadecane (1), 1-pentadecanol (2), 1-hexadecanol (2), 1-hexadecanol (1), and 1-octadecanol (2), where the number in parentheses indicates the GC-MS peak assignment method. MS fragmentation spectra of Method 1 peaks are shown in FIG. 5, plotted against their respective library hits. The major products were n-pentadecane and n-heptadecene.

The formation of these fourteen products is consistent with both the expected incomplete operation, i.e., acyl-ACP→fatty aldehyde→fatty alcohol, and expected complete operation, i.e., acyl-ACP→fatty aldehyde→alkane/alkene, of the Aar-Adm pathway in *E. coli*, whose major straight-chain acyl-ACPs include 12:0, 14:0, 16:0, 18:0, 16:1Δ9cis, and 18:1Δ11cis acyl groups (Heipieper H J (2005). Adaptation of *Escherichia coli* to Ethanol on the Level of Membrane Fatty Acid Composition. *Appl Environ Microbiol* 71:3388). Assuming that n-pentadecene (2) is derived 16:1Δ9cis-ACP, it would correspond to cis-7-pentadecene. Indeed, an n-pen-

tadecene isomer was identified as the highest-confidence MS fragmentation library hit at that retention time, with the expected molecular ion of molecular weight 210; also, as expected, it elutes slightly before n-pentadecane. With respect to 1-dodecanol (2), a sufficiently clean fragmentation spectrum could not be obtained for that peak due to the overlapping, much larger n-pentadecane (1) peak. Its presence, however, is consistent with the existence of 12:0-ACP in *E. coli*, and its retention time is exactly that extrapolated from the relationship between 1-alkanol carbon number and observed retention time, for the 1-tetradecanol, 1-hexadecanol, and 1-octadecanol authentic standards that were run. Assuming that n-hexadecene (2) is derived from the trace-level unsaturated 17:1 Δ 9cis acyl group expected in the *E. coli* acyl-ACP population due to rare acyl chain initiation with propionyl-CoA as opposed to malonyl-CoA, it would correspond to cis-8-hexadecene. Indeed, an n-hexadecene isomer was identified as the highest-confidence MS fragmentation library hit at that retention time, with the expected molecular ion of molecular weight 224; also, as expected, it elutes slightly before n-heptadecane. Assuming that n-heptadecene (2) is derived 18:1 Δ 11 cis-ACP, it would correspond to cis-7-heptadecene. Indeed, an n-heptadecene isomer was identified as the highest-confidence MS fragmentation library hit at that retention time, with the expected molecular ion of molecular weight 238; also, as expected, it elutes slightly before n-heptadecane. With respect to 1-pentadecanol (2), a sufficiently clean fragmentation spectrum could not be obtained for that peak due to its low abundance. Its presence, however, is consistent with the existence of trace-level 15:0 acyl group expected in the *E. coli* acyl-ACP population due to rare acyl chain initiation with propionyl-CoA as opposed to malonyl-CoA, and its retention time is exactly that interpolated from the relationship between 1-alkanol carbon number and observed retention time, for the 1-tetradecanol, 1-hexadecanol, and 1-octadecanol authentic standards that were run. In addition, 1-pentadecanol was identified as the highest-confidence MS fragmentation library hit at that retention time in acetone extracts of JCC1170, a BL21(DE3) derivative that expresses Aar without Adm (see Example 4). With respect to 1-hexadecanol (2), a sufficiently clean fragmentation spectrum could not be obtained for that peak due to its low abundance; however, assuming that it is derived 16:1 Δ 9cis-ACP, it would correspond to cis-9-hexadecen-1-ol. Also, as expected, it elutes slightly before 1-hexadecanol. Finally, assuming that n-octadecanol (2) is derived 18:1 Δ 9cis-ACP, it would correspond to cis-11-octadecen-1-ol. Indeed, an n-octadecen-1-ol isomer was identified as the highest-confidence MS fragmentation library hit at that retention time, with the expected molecular ion of molecular weight 250; also, as expected, it elutes slightly before 1-octadecanol.

Example 4

Production of Fatty Alcohols in *Escherichia coli* B Through Heterologous Expression of *Synechococcus elongatus* SYN-PCC7942_1594 (aar) without Co-Expression of SYN-PCC7942_1593 (adm)

In order to test the hypothesis that both AAR and ADM are required for alkane biosynthesis, as well as the prediction that expression of AAR alone should result in the production of fatty alcohols only in *E. coli* (due to non-specific dehydrogenation of the fatty aldehydes generated), expression constructs containing just SYN-PCC7942_1593 (ADM) and just SYN-PCC7942_1594 (AAR), were created. Accordingly, the SYN-PCC7942_1593 and SYN-PCC7942_1594 coding

sequences were individually PCR-amplified and cloned via NdeI and MfeI into the commercial expression vector pCDF-DuetTM-1 (Novagen). The resulting plasmids were denoted pJB881 (SYN-PCC7942_1593 only) and pJB882 (SYN-PCC7942_1594 only); in each construct, the coding sequence was placed under the transcriptional control of the inducible T7lacO promoter.

The intracellular hydrocarbon products of *E. coli* BL21 (DE3) (Novagen) harboring pJB881, JCC1169, and of *E. coli* BL21(DE3) (Novagen) harboring pJB882, JCC1170, were compared to those of *E. coli* BL21(DE3) harboring pCDF-DuetTM-1, the negative control strain JCC114, as well as to the positive control SYN-PCC7942_1593-SYN-PCC7942_1594 strain JCC1113 (Example 3), by gas chromatography-mass spectrometry (GC-MS). Clonal cultures of JCC1169, JCC1170, JCC1114, and JCC1113 were grown, extracted, and analyzed by GC-MS as described in Example 3, with the following exception: the JCC1170 culture was grown overnight in M9fs medium without IPTG, because the culture did not grow if IPTG was added. Presumably, this was due to the toxic over-accumulation of fatty alcohols that occurred even in the absence of inducer.

The TICs of JCC1169, JCC1170, JCC1114, and JCC1113 acetone cell pellet extractants are shown in FIG. 6; n-alkane and 1-alkanol standard traces have been omitted. Hydrocarbons identified in JCC1170, but not in control strain JCC1114, are detailed in Table 5.

TABLE 5

Hydrocarbons detected by GC-MS in acetone cell pellet extractants of JCC1170 but not JCC1114 in increasing order of retention time.

Compound	JCC1114	JCC1170	GC-MS Peak Assignment	Candidate isomer
1-tetradecanol	-	+	Method 1	
1-pentadecanol	-	+	Method 2 (envelope-type MS with molecular ion mass 182)	
1-hexadecanol	-	+	Method 2 (envelope-type MS with molecular ion mass 222)	cis-9-hexadecen-1-ol
1-hexadecanol	-	+	Method 1	
1-octadecanol	-	+	Method 2 (envelope-type MS with molecular ion mass 250)	cis-11-octadecen-1-ol

"-" not detected;
 "+" detected.

These hydrocarbons are 1-tetradecanol (1), 1-pentadecanol (2), 1-hexadecanol (2), 1-hexadecanol (1), and 1-octadecanol (2), where the number in parentheses indicates the GC-MS peak assignment method. MS fragmentation spectra of Method 1 peaks are shown in FIG. 7, plotted against their respective library hits. No hydrocarbons were identified in JCC1169, whose trace was indistinguishable from that of JCC1114, as expected owing to absence of fatty aldehyde substrate generation by AAR.

The lack of production of alkanes, alkenes, and fatty alkanols in JCC1169, the production of only fatty alcohols in JCC1170, and the production of alkanes, alkenes, and fatty alkanols in JCC1113 (as discussed in Example 3) are all consistent with the proposed mechanism of alkane biosynthesis by AAR and ADM in *E. coli*. Thus, the formation of the

five fatty alcohols in JCC1170 is consistent with only AAR being active, and active on the known straight-chain acyl-ACPs (see Example 3). With respect to 1-pentadecanol (2), its presence is consistent with the existence of trace-level 15:0 acyl group expected in the *E. coli* acyl-ACP population due to rare acyl chain initiation with propionyl-CoA as opposed to malonyl-CoA and its retention time is exactly that interpolated from the relationship between 1-alkanol carbon number and observed retention time, for the 1-tetradecanol, 1-hexadecanol, and 1-octadecanol authentic standards that were run. Most importantly, the 1-pentadecanol (2) peak exhibits an envelope-type fragmentation mass spectrum, with the expected molecular ion of molecular weight 182. Unlike in the case of JCC1113, a clean fragmentation spectrum from the candidate 1-hexadecanol peak could now be obtained due to increased abundance. The top library hit was a 1-hexadecanol with the expected molecular ion of molecular weight 222. Assuming that it is derived from 16:1 Δ 9cis hexadecenyl-ACP, the isomeric assignment would be cis-9-hexadecen-1-ol; also, as expected, it elutes slightly before 1-hexadecanol. Assuming that n-octadecanol (2) is derived from 18:1 Δ 9cis-ACP, it would correspond to cis-11-octadecen-1-ol. Indeed, an n-octadecen-1-ol isomer was identified as the highest-confidence MS fragmentation library hit at that retention time, with the expected molecular ion of molecular weight 250; also, as expected, it elutes slightly before 1-octadecanol. There is also an unidentified side peak in JCC1170 that elutes in the tail of 1-hexadecanol and whose fragmentation mass spectrum was not sufficiently clean to enable possible identification. It is hypothesized that this could be the primary C₁₈ aldehyde product expected of AAR-only activity in *E. coli*, i.e., cis-11-octadecenal.

Example 5

Production of n-Alkanes, n-Alkenes, and Fatty Alcohol in *Synechococcus* sp. PCC 7002 Through Heterologous Expression of *Synechococcus elongatus* PCC7942 SYN-PCC7942_1593 (adm) and SYN-PCC7942_1594 (aar)

In order to test whether heterologous expression of AAR and ADM would lead to the desired alkane biosynthesis in a cyanobacterial host, the SYN-PCC7942_1593-SYN-PCC7942_594 operon was expressed in *Synechococcus* sp. PCC 7002 (JCC138). Accordingly, plasmid pJB823 was transformed into JCC138, generating strain JCC1160. The sequence and annotation of this plasmid is provided as SEQ ID NO: 15, and described in Example 2. In this construct, the SYN-PCC7942_1593-SYN-PCC7942_594 operon is placed under the transcriptional control of the constitutive aphII promoter. 500 base pair upstream and downstream homology regions direct homologous recombinational integration into the native high-copy pAQ1 plasmid of JCC138, and an *aadA* gene permits selection of transformants by virtue of their resistance to spectinomycin.

To test the effect of potentially stronger promoters, constructs directly analogous to pJB823 were also generated that substituted the aphII promoter with the following: the promoter of *cro* from lambda phage (PcI), the promoter of *cpcB* from *Synechocystis* sp. PCC 6803 (PcpcB), the *trc* promoter along with an upstream copy of a promoter-lacI cassette (PlacI-trc), the synthetic EM7 promoter (PEM7). Promoters were exchanged via the NotI and NdeI sites flanking the promoter upstream of the SYN-PCC7942_1593-SYN-

PCC7942_594 operon in the pJB823 vector. The corresponding final plasmids were as follows: pJB886 (PcI), pJB887 (PcpcB), pJB889 (PlacI-trc), pJB888 (PEM7), and pJB823 (PaphII). These sequences of pJB886, pJB887, pJB889, and pJB888 are identical to the sequence of pJB823 except in the region between the NotI and NdeI sites, where they differ according to the promoter used. The sequences of the different promoter regions are provided as SEQ ID NO: 19 (PO, SEQ ID NO: 20 (PcpcB), SEQ ID NO: 21 (PlacI-trc), and SEQ ID NO: 22 (PEM7). The sequence of the PaphII promoter is presented in SEQ ID NO: 15.

pJB886, pJB887, pJB889, pJB888, pJB823, as well as pJB5 (the empty pAQ1 targeting vector that entirely lacked the SYN-PCC7942_1593-SYN-PCC7942_594 operonic sequence) were naturally transformed into JCC138 using a standard cyanobacterial transformation protocol, generating strains JCC1221 (PcI), JCC1220 (PcpcB), JCC1160b (PlacI-trc), JCC1160a (PEM7), JCC1160 (PaphII), and JCC879 (pJB5), respectively. Briefly, 5-10 μ g of plasmid DNA was added to 1 ml of neat JCC138 culture that had been grown to an OD₇₃₀ of approximately 1.0. The cell-DNA mixture was incubated at 37° C. for 4 hours in the dark with gentle mixing, plated onto A+ plates, and incubated in a photoincubator (Percival) for 24 hours, at which point spectinomycin was underlaid to a final concentration of 50 μ g/ml. Spectinomycin-resistant colonies appeared after 5-8 days of further incubation under 24 hr-light conditions (~100 μ mol photons m⁻² s⁻¹). Following one round of colony purification on A+ plates supplemented with 100 μ g/ml spectinomycin, single colonies of each of the six transformed strains were grown in test-tubes for 4-8 days at 37° C. at 150 rpm in 3% CO₂-enriched air at ~100 μ mol photons m⁻² s⁻¹ in a Multitron II (Infors) shaking photoincubator. The growth medium used for liquid culture was A+ with 200 μ g/ml spectinomycin.

In order to compare the intracellular hydrocarbon products of strains JCC1221, JCC1220, JCC1160b, JCC1160a, JCC1160, and JCC879, 24 OD₇₃₀-ml worth of cells (~2.4 \times 10⁹ cells) of each strain was collected from the aforementioned test-tube cultures by centrifugation. Cell pellets were washed thoroughly by 3 cycles of resuspension in Milli-Q water and microcentrifugation, and then dewetted as much as possible by 3 cycles of microcentrifugation and aspiration. Cell pellets were then extracted by vortexing for 5 minutes in 0.7 ml acetone containing 20 μ g/ml BHT and 20 μ g/ml EA. Cell debris was pelleted by centrifugation, and 600 μ l supernatant was pipetted into a GC vial. The six extractants, along with authentic standards, were then analyzed by GC-MS as described in Example 2.

The TICs of JCC1221, JCC1220, JCC1160b, JCC1160a, JCC1160, and JCC879 acetone cell pellet extractants are shown in FIG. 8; n-alkane and 1-alkanol standards are as in Example 2. Consistent with a range of promoter strengths, and with function of the AAR-ADM pathway, there was a range of hydrocarbon accumulation, the order of accumulation being PcI>PcpcB>PlacI-trc>PEM7>PaphII (FIG. 8A).

In JCC1160, approximately 0.2% of dry cell weight was found as n-alkanes and n-alkan-1-ol (excluding n-nonadec-1-ene). Of this 0.2%, approximately three-quarters corresponded to n-alkanes, primary products being n-heptadecane and n-pentadecane. These hydrocarbons were not detected in JCC879. The data are summarized in Table 6A.

47

TABLE 6A

Hydrocarbons detected in acetone extracts of JCC1160 and JCC879.		
Compound	Approximate % of dry cell weight	
	JCC879	JCC1160
n-pentadecane	not detected	0.024%
n-hexadecane	nd	0.004%
n-heptadecane	nd	0.110%
n-octadecan-1-ol	nd	0.043%
	Total	0.181%
	% of products that are n-alkanes	76%

The highest accumulator was JCC1221 (PcI). Hydrocarbons identified in JCC1221, but not in control strain JCC879, are detailed in Table 6B, Table 6C and FIG. 8B. These hydrocarbons are n-tridecane (1), n-tetradecane (1), n-pentadecene (2), n-pentadecane (1), n-hexadecane (1), n-heptadec-di-ene (2), three isomers of n-heptadecene (2), n-heptadecane (1), and 1-octadecanol (1), where the number in parentheses indicates the GC-MS peak assignment method.

TABLE 6B

n-Alkanes quantitated in acetone extract of JCC1221	
Compound	% of JCC1221 dry cell weight
n-tridecane	<0.001%
n-tetradecane	0.0064%
n-pentadecane	0.40%
n-hexadecane	0.040%
n-heptadecane	1.2%
Total	1.67%

MS fragmentation spectra of Method 1 peaks are shown in FIG. 9, plotted against their respective library hits. The only alkanes/alkenes observed in JCC879 were 1-nonadecene and a smaller amount of nonadec-di-ene, alkenes that are known to be naturally synthesized by JCC138 (Winters K et al. (1969) Science 163:467-468). The major products observed in JCC1221 were n-pentadecane (~25%) and n-heptadecane (~75%); all others were in relatively trace levels.

The formation of n-pentadecane and n-heptadecane in JCC1221, as well as the nine other trace hydrocarbon products, is consistent with the virtually complete operation of the ADM-AAR pathway in JCC138, i.e., 16:0 hexadecyl-ACP→n-hexadecanal→n-pentadecane and 18:0 octadecyl-ACP→n-octadecanal→n-heptadecane. Indeed it is known that the major acyl-ACP species in this organism are C_{16:0} and C_{18:0} (Murata N et al. (1992) Plant Cell Physiol 33:933-941). Relatively much less fatty alcohol is produced relative to AAR-ADM expression in *E. coli* (Example 3), as expected given the presence in JCC138 of a cyanobacterial ferredoxin/ferredoxin-NADPH reductase system that can regenerate the di-iron active site of ADM, thereby preventing the accumulation of hexadecanal and octadecanal that could in turn be non-specifically dehydrogenated to the corresponding 1-alkanols. Thus, in JCC1221, only a very small 1-octadecanol (1) peak is observed (FIG. 8).

The other trace hydrocarbons seen in JCC1221 are believed to be unsaturated isomers of n-pentadecane and n-heptadecane (Table 6C). It is hypothesized that all these alkenes are generated by desaturation events following the production of the corresponding alkanes by the SYN-PCC7942_1593 Adm. This contrasts with the situation in *E.*

48

coli, where double bonds are introduced into the growing acyl chain while it is linked to the acyl carrier protein (Example 3). JCC138 is known to have a variety of position-specific acyl-lipid desaturases that, while nominally active only on fatty acids esterified to glycerolipids, could potentially act on otherwise unreactive alkanes produced nonphysiologically by the action of AAR and ADM. JCC138 desaturases, i.e., DesA, DesB, and DesC, introduce cis double bonds at the 49, 412, and 415 positions of C₁₈ acyl chains, and at the Δ9 and Δ12 positions of C₁₆ acyl chains (Murata N and Wada H (1995) Biochem J. 308:1-8). The candidate n-pentadecene peak is believed to be cis-4-pentadecene (Table 6C).

Assuming also that heptadecane could also serve as a substrate for JCC138 desaturases, and that it would be desaturated at positions analogous to the Δ9, Δ12, and Δ15 of the C₁₈ acyl moiety, there are four theoretically possible mono-unsaturated isomers: cis-3-heptadecene, cis-6-heptadecene, cis-8-heptadecene, and cis-9-heptadecene. These isomers do not include the single n-heptadecene species nominally observed in *E. coli*, cis-7-heptadecene (Example 2). It is believed that the three peaks closest to the n-heptadecane peak—denoted by subscripts 1, 2, and 3 in Table 6C and FIG. 8B—encompass at least three of these four mono-unsaturated heptadecane isomers. Consistent with this, n-heptadecene₂ and n-heptadecene₃ peaks have the expected molecular ions of mass 238 in their envelope-type fragmentation spectra. There are many isomeric possibilities, accordingly, for the putative cis,cis-heptadec-di-ene peak, which has an envelope-type fragmentation spectrum with the expected molecular ions of mass 236. As expected, all putative heptadecene species elute slightly before n-heptadecane.

TABLE 6C

Alkane and alkenes detected by GC-MS in acetone cell pellet extractants of JCC1221 but not JCC879 in increasing order of retention time.				
Compound	JCC879	JCC1221	GC-MS Assignment	Candidate isomer
n-tridecane	-	+	Method 1	
n-tetradecane	-	+	Method 1	
n-pentadecene	-	+	Method 2	cis-4-pentadecene
n-pentadecane	-	+	Method 1	
n-hexadecane	-	+	Method 1	
n-heptadec-di-ene	-	+	Method 2 (envelope-type MS with molecular ion mass 236)	cis,cis-heptadec-di-ene
n-heptadecene ₃	-	+	Method 2 (envelope-type MS with molecular ion mass 238)	cis-[3/6/8/9]-heptadecene
n-heptadecene ₂	-	+	Method 2 (envelope-type MS with molecular ion mass 238)	cis-[3/6/8/9]-heptadecene
n-heptadecene ₁	-	+	Method 2	cis-[3/6/8/9]-heptadecene
n-heptadecane	-	+	Method 1	
1-octadecanol	-	+	Method 1	

65 “-” not detected;
“+” detected.

Example 6

Intracellular Accumulation of n-Alkanes to Up to 5% of Dry Cell Weight in *Synechococcus* sp. PCC 7002 Through Heterologous Expression of *Synechococcus elongatus* PCC7942 SYN-PCC7942_1593 (adm) and SYN-PCC7942_594 (aar)

In order to quantitate more accurately the level of intracellular accumulation of n-alkane products in the alkanogen JCC1221 (Example 5), the levels of n-pentadecane and n-heptadecane, as well as the relatively trace products n-tetradecane and n-hexadecane, were quantified with respect to dry cell weight (DCW). Based on the hypothesis that the extent of n-alkane production could correlate positively with the level of SYN-PCC7942_1593-SYN-PCC7942_594 operon expression, the DCW-normalized n-alkane levels of JCC1221 were determined as a function of the spectinomycin concentration of the growth medium. The rationale was that the higher the spectinomycin selective pressure, the higher the relative copy number of pAQ1, and the more copies of the aadA-linked SYN-PCC7942_1593-SYN-PCC7942_594 operon.

A clonal starter culture of JCC1221 was grown up in A+ medium supplemented with 100 µg/ml spectinomycin in for 7 days at 37° C. at 150 rpm in 3% CO₂-enriched air at ~100 µmol photons m⁻² s⁻¹ in a Multitron II (Infors) shaking photoincubator. At this point, this culture was used to inoculate triplicate 30 ml JB2.1 medium (PCT/US2009/006516) flask cultures supplemented with 100, 200, 300, 400, or 600 µg/ml spectinomycin. JB2.1 medium consists of 18.0 g/l sodium chloride, 5.0 g/l magnesium sulfate heptahydrate, 4.0 g/l sodium nitrate, 1.0 g/l Tris, 0.6 g/l potassium chloride, 0.3 g/l calcium chloride (anhydrous), 0.2 g/l potassium phosphate monobasic, 34.3 mg/l boric acid, 29.4 mg/l EDTA (disodium salt dihydrate), 14.1 mg/l iron (III) citrate hydrate, 4.3 mg/l manganese chloride tetrahydrate, 315.0 µg/l zinc chloride, 30.0 µg/l molybdenum (VI) oxide, 12.2 µg/l cobalt (II) chloride hexahydrate, 10.0 µg/l vitamin B₁₂, and 3.0 µg/l copper (II) sulfate pentahydrate. The 15 cultures were grown for 10 days at 37° C. at 150 rpm in 3% CO₂-enriched air at ~100 µmol photons m⁻² s⁻¹ in a Multitron II (Infors) shaking photoincubator.

For each culture, 5-10 ml was used for dry cell weight determination. To do so, a defined volume of culture—corresponding to approximately 20 mg DCW—was centrifuged to pellet the cells. Cells were transferred to a pre-weighed eppendorf tube, and then washed by 2 cycles of resuspension in Milli-Q water and microcentrifugation, and dewetted by 3 cycles of microcentrifugation and aspiration. Wet cell pellets were frozen at -80° C. for two hours and then lyophilized overnight, at which point the tube containing the dry cell mass was weighed again such that the mass of the cell pellet could be calculated within ±0.1 mg. In addition, for each culture, 0.3-0.8 ml was used for acetone extraction of the cell pellet for GC analysis. To do so, a defined volume of culture—corresponding to approximately 1.4 mg DCW—was microcentrifuged to pellet the cells. Cells were then washed by 2 cycles of resuspension in Milli-Q water and microcentrifugation, and then dewetted by 4 cycles of microcentrifugation and aspiration. Dewetted cell pellets were then extracted by vortexing for 1 minute in 1.0 ml acetone containing 50 µg/ml BHT and 160 µg/ml n-heptacosane internal standard (Sigma 51559). Cell debris was pelleted by centrifugation, and 700 µl supernatant was pipetted into a GC vial.

Concentrations of n-tetradecane, n-pentadecane, n-hexadecane, and n-heptadecane in the fifteen extractants were

quantitated by gas chromatography/flame ionization detection (GC/FID). Unknown n-alkane peak areas in biological samples were converted to concentrations via linear calibration relationships determined between known n-tetradecane, n-pentadecane, n-hexadecane, and n-heptadecane authentic standard concentrations and their corresponding GC-FID peak areas. Standards were obtained from Sigma. GC-FID conditions were as follows. An Agilent 7890A GC/FID equipped with a 7683 series autosampler was used. 1 µl of each sample was injected into the GC inlet (split 5:1, pressure: 20 psi, pulse time: 0.3 min, purge time: 0.2 min, purge flow: 15 ml/min) and an inlet temperature of 280° C. The column was a HP-5MS (Agilent, 30 m×0.25 mm×0.25 µm) and the carrier gas was helium at a flow of 1.0 ml/min. The GC oven temperature program was 50° C., hold one minute; 10° C./min increase to 280° C.; hold ten minutes. n-Alkane production was calculated as a percentage of the DCW extracted in acetone.

Consistent with scaling between pAQ1 selective pressure and the extent of intracellular n-alkane production in JCC1221, there was a roughly positive relationship between the % n-alkanes with respect to DCW and spectinomycin concentration (FIG. 10). For all JCC1221 cultures, n-alkanes were ~25% n-pentadecane and ~75% n-heptadecane. The minimum n-alkane production was ~1.8% of DCW at 100 µg/ml spectinomycin and 5.0% in one of the 600 µg/ml spectinomycin cultures.

Example 7

Production of n-Alkanes in *Synechococcus* sp. PCC 7002 Through Heterologous Expression of *Prochlorococcus Marinus* MIT 9312 PMT9312_0532 (adm) and PMT9312_0533 (aar)

This candidate Adm/Aar pair from *Prochlorococcus marinus* MIT9312 was selected for functional testing by heterologous expression in JCC138 because of the relatively low amino acid homology (≤62%) of these proteins to their *Synechococcus elongatus* PCC7942 counterparts, SYN-PCC7942_1593 and SYN-PCC7942_594. Specifically, the 252-amino acid protein PMT9312_0532 exhibits only 62% amino acid identity with the 232 amino acid protein SYN-PCC7942_1593, wherein amino acids 33-246 of the former are aligned with amino acids 11-224 of the latter. The 347 amino acid protein PMT9312_0533 exhibits only 61% amino acid identity with the 342 amino acid protein SYN-PCC7942_594, wherein amino acids 1-337 of the former are aligned with amino acids 1-339 of the latter.

A codon- and restriction-site-optimized version of the PMT9312_0532-PMT9312_0533 operon was synthesized by DNA2.0 (Menlo Park, Calif.), flanked by NdeI and EcoRI sites. The operon was cloned into the pAQ1 homologous recombination vector pJB5 via NdeI and EcoRI, such that the PMT9312_0532-PMT9312_0533 operon was placed under transcriptional control of the aphII promoter. The sequence of the pJB947 vector is provided as SEQ ID NO: 17.

pJB947 was transformed into JCC138 as described in Example 5, generating strain JCC1281. The hydrocarbon products of this strain were compared to those of the negative control strain JCC879, corresponding to JCC138 transformed with empty pJB5 (see Example 5). Eight OD₇₃₀-ml worth of cells (~8×10⁸ cells) of each strain was collected by centrifugation, having been grown in A+ medium supplemented with 200 µg/ml spectinomycin as described in Example 5. Cell pellets were washed thoroughly by 3 cycles of resuspension in Milli-Q water and microcentrifugation, and then dewetted

51

as much as possible by 3 cycles of microcentrifugation and aspiration. Cell pellets were then extracted by vortexing for 5 minutes in 0.7 ml acetone containing 20 µg/ml BHT and 20 µg/ml EA. Cell debris was pelleted by centrifugation, and 600 µl supernatant was pipetted into a GC vial. Samples were analyzed by GC-MS as described in Example 5.

The TICs of JCC1281 and JCC879 acetone cell pellet extractants are shown in FIG. 11; n-alkane standards are as in Example 6. Hydrocarbons identified in JCC1281, but not in control strain JCC879, were n-pentadecane (1) and n-heptadecane (1), where the number in parentheses indicates the GC-MS peak assignment method. MS fragmentation spectra of Method 1 peaks are shown in FIG. 12, plotted against their respective library hits (as noted in Example 5, the only alkanes/alkenes observed in JCC879 were 1-nonadecene and a smaller amount of nonadec-di-ene, alkenes that are known to be naturally synthesized by JCC138). The amount of n-alkanes produced in JCC1281 is at least 0.1% dry cell weight, and at least 2-two times higher than the amount produced by JCC879. The ratio of n-pentadecane:n-heptadecane (~40%:~60%) in JCC1281 was higher than that observed in JCC1221 (~25%:~75%), suggesting that the PMT9312_0532 (ADM) and/or the PMT9312_0533 (AAR) exhibit higher activity towards the C₁₆ substrates relative to C₁₈ substrates, compared to SYN-PCC7942_1593 (ADM) and/or SYN-PCC7942_594 (AAR).

Example 8

Augmentation of Native n-Alkane Production in *Thermosynechococcus Elongatus* BP-1 by Overexpression of the Native tll1313 (adm)-tll1312 (aar) Operon

Genes encoding *Thermosynechococcus elongatus* BP-1 tll1312 (AAR) and tll1313 (ADM) are incorporated into one or more plasmids (e.g., pJB5 derivatives), comprising promoters of differing strength. The plasmids are used to transform *Thermosynechococcus elongatus* BP-1. Overexpression of the genes in the transformed cells are measured as will the amount of n-alkanes, particularly heptadecane, produced by the transformed cells, in a manner similar to that described in Example 3. The n-alkanes and other carbon-based products of interest can also be isolated from the cell or cell culture, as needed.

Wild-type *Thermosynechococcus elongatus* BP-1, referred to as JCC3, naturally produces n-heptadecane as the major intracellular hydrocarbon product, with traces of n-hexadecane and n-pentadecane. These n-alkanes were identified by GC-MS using Method 1; fragmentation spectra are shown in FIG. 13. Briefly, a colony of JCC3 was grown in B-HEPES medium to a final OD₇₃₀ of ~4, at which point 5 OD₇₃₀-ml worth of cells was harvested, extracted in acetone, and analyzed by GC-MS as detailed in Example 5.

In an effort to augment this n-alkane production, the native tll1313-tll1312 operonic sequence from this organism was PCR-amplified and cloned into the *Thermosynechococcus elongatus* BP-1 chromosomal integration vector pJB825. This construct places the tll1313-tll1312 operon under the transcriptional control of the constitutive a promoter. The sequence of the resulting plasmid, pJB825t, is shown in SEQ ID NO:18.

pJB825 and pJB825t were naturally transformed into JCC3 using a standard cyanobacterial transformation protocol, generating strains JCC1084 and JCC1084t, respectively. Briefly, 25 µg of plasmid DNA was added to 0.5 ml of concentrated JCC3 culture (OD₇₃₀ ~100) that had originally been grown to

52

an OD₇₃₀ of approximately 1.0 in B-HEPES at 45° C. in 3% CO₂-enriched air at ~100 µmol photons m⁻² s⁻¹ in a Multitron II (Infors) shaking photoincubator. The cell-DNA mixture was incubated at 37° C. for 4 hours in the dark with gentle mixing, made up to 7 ml with fresh B-HEPES medium, and then incubated under continuous light conditions (~100 µmol photons m⁻² s⁻¹) for 20 hours at 45° C. at 150 rpm in 3% CO₂-enriched air at ~100 µmol photons m⁻² s⁻¹ in a Multitron II (Infors) shaking photoincubator. At this point, cells were collected by centrifugation and serial dilutions were mixed with molten top agar and plated on the surface of B-HEPES plates supplemented with 60 µg/ml kanamycin. Transformant colonies appeared in the top agar layer within around 7 days upon incubation in a photoincubator (Percival) in 1% CO₂-enriched air at continuous ~100 µmol photons m⁻² s⁻¹ irradiance. Single colonies of JCC1084 and JCC1084t were then grown up in triplicate to an OD₇₃₀ of ~6 in B-HEPES/60 µg/ml kanamycin liquid culture, and their intracellular hydrocarbon products quantitated by GC-FID.

3.5 OD₇₃₀-ml worth of cells (~3.5×10⁸ cells) of each replicate culture of each strain was collected by centrifugation. Cell pellets were washed thoroughly by 3 cycles of resuspension in Milli-Q water and microcentrifugation, and then dewatered as much as possible by 3 cycles of microcentrifugation and aspiration. Cell pellets were then extracted by vortexing for 1 minutes in 0.7 ml acetone containing 20 µg/ml BHT and 20 µg/ml n-heptacosane. Cell debris was pelleted by centrifugation, and 600 µl supernatant was pipetted into a GC vial. The two extractants, along with authentic C₈-C₂₀ n-alkane authentic standards (Sigma 04070), were then analyzed by GC coupled with flame ionization detection (FID) as described in Example 6. Quantitation of n-pentadecane, n-hexadecane, and n-heptadecane by GC-FID, and dry cell weights were taken as described in Example 6.

Consistent with increased expression of tll1313-tll1312 in JCC1084t relative to the control strain JCC1084, n-pentadecane, n-hexadecane, and n-heptadecane were ~500%, ~100%, and ~100% higher, respectively, in JCC1084t relative to their % DCW levels in JCC1084 (FIG. 14). The total n-alkane concentration in both strains was less than 1%. The n-alkane concentration in JCC1084t was at least 0.62% and at least twice as much n-alkane was produced relative to JCC1084.

Example 9

Comparison of Intracellular Hydrocarbon Products of JCC1113 (a Derivative of *E. Coli*) and JCC1221 (a Derivative of *Synechococcus* sp. PCC 7002), Both Strains Heterologously Expressing *Synechococcus elongatus* SYN-PCC7942_1593 (adm) and SYN-PCC7942_1594 (aar)

GC-MS TICs of JCC1113 and JCC1221 acetone cell pellet extractants are shown in FIG. 15, along with the TIC of C₈-C₂₀ n-alkane authentic standards (Sigma 04070). These two strains are derived from *E. coli* BL21(DE3) and *Synechococcus* sp. PCC7002, respectively, and are described in detail in Examples 3 and 5, respectively. JCC1113 synthesizes predominantly n-heptadecene and n-pentadecane, whereas JCC1221 synthesizes predominantly n-heptadecane and n-pentadecane. This figure visually emphasizes the different retention times of the n-heptadecene isomer produced in JCC1113 and n-heptadecane produced in JCC1221.

Example 10

Production of Hydrocarbons in Yeast

The methods of the invention can be performed in a number of lower eukaryotes such as *Saccharomyces cerevisiae*, *Trichoderma reesei*, *Aspergillus nidulans* and *Pichia pastoris*. Engineering such organisms may include optimization of genes for efficient transcription and/or translation of the encoded protein. For instance, because the ADM and AAR genes introduced into a fungal host are of cyanobacterial origin, it may be necessary to optimize the base pair composition. This includes codon optimization to ensure that the cellular pools of tRNA are sufficient. The foreign genes (ORFs) may contain motifs detrimental to complete transcription/translation in the fungal host and, thus, may require substitution to more amenable sequences. The expression of each introduced protein can be followed both at the transcriptional and translational stages by well known Northern and Western blotting techniques, respectively.

Use of various yeast expression vectors including genes encoding activities which promote the ADM or AAR pathways, a promoter, a terminator, a selectable marker and targeting flanking regions. Such promoters, terminators, selectable markers and flanking regions are readily available in the art. In a preferred embodiment, the promoter in each case is selected to provide optimal expression of the protein encoded by that particular ORF to allow sufficient catalysis of the desired enzymatic reaction. This step requires choosing a promoter that is either constitutive or inducible, and provides regulated levels of transcription. In another embodiment, the terminator selected enables sufficient termination of transcription. In yet another embodiment, the selectable/counterselectable markers used are unique to each ORF to enable the subsequent selection of a fungal strain that contains a specific combination of the ORFs to be introduced. In a further embodiment, the locus to which relevant plasmid construct (encoding promoter, ORF and terminator) is localized, is determined by the choice of flanking region.

The engineered strains can be transformed with a range of different genes for production of carbon-based products of interest, and these genes are stably integrated to ensure that the desired activity is maintained throughout the fermentation process. Various combinations of enzyme activities can be engineered into the fungal host such as the ADM, ADR pathways while undesired pathways are attenuated or knocked out.

Example 11

Quantitation of Intracellular
n-Pentadecane:n-Heptadecane Ratio of
Synechococcus sp. PCC 7002 Strains Constitutively
Expressing Heterologous *Synechococcus elongatus*
SYNPCC7942_1593 (adm) Plus
SYNPCC7942_1594 (aar) or Heterologous
Prochlorococcus marinus MIT 9312
PMT9312_0532 (adm) Plus PMT9312_0533 (aar)
on pAQ1

In Example 5 ("Production of n-Alkanes, n-Alkenes, and Fatty Alcohol in *Synechococcus* sp. PCC 7002 through Heterologous Expression of *Synechococcus elongatus* PCC7942 SYNPCC7942_1593 (adm) and SYNPCC7942_1594 (aar)") and Example 7 ("Production of n-Alkanes in *Synechococcus* sp. PCC 7002 through Heterologous Expression of *Prochlorococcus marinus* MIT 9312 PMT9312_0532 (adm)

and PMT9312_0533 (aar)"), the intracellular hydrocarbon products of JCC138 (*Synechococcus* sp. PCC 7002) strains expressing the *Synechococcus elongatus* sp. PCC7942 and *Prochlorococcus marinus* MIT 9312 adm-aar operons were analyzed by GC-MS. In this Example, GC-FID (Gas Chromatography-Flame Ionization Detection) was applied to more accurately measure these products with respect to dry cell weight. Of special interest was the ratio between n-pentadecane and n-heptadecane. In this regard, it is noted that *Synechococcus elongatus* sp. PCC7942 naturally synthesizes n-heptadecane as the major intracellular n-alkane, whereas *Prochlorococcus marinus* MIT 9312 naturally synthesizes n-pentadecane as the major intracellular n-alkane.

The following four strains were compared: (1) JCC138, corresponding to wild-type *Synechococcus* sp. PCC 7002, (2) JCC879, corresponding to negative control strain JCC138 transformed with pAQ1-targeting plasmid pJB5 described in Example 5, (3) JCC1469, corresponding to JCC138 Δ SYNPCC7002_A1173::gent (JCC1218) transformed with pAQ1-targeting plasmid pJB886 encoding constitutively expressed *Synechococcus elongatus* sp. PCC7942 adm-aar described in Example 5, and (4) JCC1281, corresponding to JCC138 transformed with pAQ1-targeting plasmid pJB947 encoding constitutively expressed *Prochlorococcus marinus* MIT 9312 adm-aar, described in Example 7. A clonal starter culture of each strain was grown up for 5 days at 37° C. at 150 rpm in 2% CO₂-enriched air at ~100 μ mol photons m⁻² s⁻¹ in a Multitron II (Infors) shaking photoincubator in A+ (JCC138), A+ supplemented with 100 μ g/ml spectinomycin (JCC879 and JCC1281), or A+ supplemented with 100 μ g/ml spectinomycin and 50 μ g/ml gentamycin (JCC1469). At this point, each starter culture was used to inoculate duplicate 30 ml JB2.1 medium flask cultures supplemented with no antibiotics (JCC138) or 400 μ g/ml spectinomycin (JCC879, JCC1469, and JCC1281). The eight cultures were then grown for 14 days at 37° C. at 150 rpm in 2% CO₂-enriched air at ~100 μ mol photons m⁻² s⁻¹ in a Multitron II (Infors) shaking photoincubator.

For each culture, 25 OD₇₃₀-ml worth of cells was collected by centrifugation in a pre-weighed eppendorf tube. Cells were washed by two cycles of resuspension in Milli-Q water and microcentrifugation, and dewetted by two cycles of microcentrifugation and aspiration. Wet cell pellets were frozen at -80° C. for two hours and then lyophilized overnight, at which point the tube containing the dry cell mass was weighed again such that the mass of the cell pellet (~6 mg) could be calculated within \pm 0.1 mg. In parallel, 4 OD₇₃₀-ml worth of cells from each culture was collected by centrifugation in an eppendorf tube, washed thoroughly by three cycles of resuspension in Milli-Q water and microcentrifugation, and then dewetted as much as possible by three cycles of microcentrifugation and aspiration. Dewetted cell pellets were then extracted by vortexing for 15 seconds in 1 ml acetone containing 23.6 mg/l BHT and 24.4 mg/l n-heptacosane (C₂₇) internal standard (ABH); cell debris was pelleted by centrifugation, and 450 μ l supernatant was submitted for GC-FID. Acetone-extracted DCW was calculated as 4/25, or 16%, of the DCW measured for 25 OD₇₃₀-ml worth of cells. In parallel with the eight biological sample extractions, six empty eppendorf tubes were extracted with ABH in the same fashion. The extraction/injection efficiency of all ABH extractants was assessed by calculating the ratio between the n-heptacosane GC-FID peak area of the sample and the average n-heptacosane GC-FID peak area of the six empty-tube controls—only ratios of 100% \pm 3% were accepted (Table 7).

Concentrations of n-tridecane (C₁₃), n-tetradecane (C₁₄), n-pentadecane (C₁₅), n-hexadecane (C₁₆), n-heptadecane

(C₁₇), and n-octadecane (C₁₈), in the eight extractants were quantitated by (GC/FID). Unknown n-alkane peak areas in biological samples were converted to concentrations via linear calibration relationships determined between known n-tridecane, n-tetradecane, n-pentadecane, n-hexadecane, n-heptadecane, and n-octadecane authentic standard concentrations and their corresponding GC-FID peak areas. Based on these linear-regression calibration relationships, 95% confidence intervals (95% CI) were calculated for interpolated n-alkane concentrations in the biological samples; interpolation was used in all cases, never extrapolation. 95% confidence intervals were reported as percentages—95% CI % in Table 1—of the interpolated concentration in question. GC-FID conditions were as follows. An Agilent 7890A GC/FID equipped with a 7683 series autosampler was used. 1 µl of each sample was injected into the GC inlet (split 8:1, pressure) and an inlet temperature of 290° C. The column was a HP-5MS (Agilent, 20 m×0.18 mm×0.18 µm) and the carrier gas was helium at a flow of 1.0 ml/min. The GC oven temperature program was 80° C., hold 0.3 minutes; 17.6° C./min increase to 290° C.; hold 6 minutes. n-Alkane production was expressed as a percentage of the acetone-extracted DCW. The coefficient of variation of the n-heptacosane GC-FID peak area of the six empty-tube controls was 1.0%.

GC-FID data are summarized in Table 7. As expected, control strains JCC138 and JCC879 made no n-alkanes, whereas JCC1469 and JCC1281 made n-alkanes, ~98% of which comprised n-pentadecane and n-heptadecane. JCC1469 made significantly more n-alkanes as a percentage of DCW (~1.9%) compared to JCC1281 (~0.7%), likely explaining the relatively low final OD₇₃₀ of the JCC1469 cultures. For the duplicate JCC121 cultures expressing *Synechococcus elongatus* sp. PCC7942 adm-aar, the percentage by mass of n-pentadecane relative to n-pentadecane plus n-heptadecane was 26.2% and 25.3%, whereas it was 57.4% and 57.2% for the duplicate JCC1221 cultures expressing *Prochlorococcus marinus* MIT 9312 adm-aar (Table 7). This result quantitatively confirms that these two different adm-aar operons generate different n-alkane product length distributions when expressed in vivo in a cyanobacterial host.

TABLE 7

Strain	OD ₇₃₀	C ₂₇ - normalized extraction/ injection efficiency	C ₁₅ as % of DCW (95% CI%)	C ₁₇ as % of DCW (95% CI%)	(C ₁₅ + C ₁₇)/(C ₁₃ + C ₁₄ + C ₁₅ + C ₁₆ + C ₁₇) Mass %	C ₁₅ / (C ₁₅ + C ₁₇) Mass %
JCC138 #1	12.5	98%	nd	nd	na	na
JCC138 #2	13.5	99%	nd	nd	na	na
JCC879 #1	9.8	100%	nd	nd	na	na
JCC879 #2	8.5	101%	nd	nd	na	na
JCC1469 #1	3.1	101%	0.60% (1.1%)	1.69% (0.7%)	97.8%	26.2%
JCC1469 #2	3.2	102%	0.36% (1.0%)	1.05% (1.1%)	98.0%	25.3%
JCC1281 #1	9.7	101%	0.26% (1.2%)	0.19% (0.9%)	97.2%	57.4%
JCC1281 #2	4.8	101%	0.51% (1.9%)	0.38% (1.1%)	97.2%	57.2%

n-Pentadecane and n-heptadecane quantitated by GC-FID in acetone cell pellet extractants of JCC138, JCC879, JCC1469, and JCC1281.

n-Octadecane was not detected in any of the samples;

nd: not detected,

na: not applicable.

Quantitation of Intracellular n-Pentadecane:n-Heptadecane Ratio of *Synechococcus* sp. PCC 7002 Strains Inducibly Expressing Chromosomally-Integrated Heterologous *Prochlorococcus marinus* MIT 9312 PMT9312_0532 (adm) Plus PMT9312_0533 (aar) with or without Heterologous *Cyanothece* sp. ATCC 51142 Cce_0778 (adm) Plus Cce_1430 (aar)

In order to confirm that heterologous expression of Aar and Adm from the chromosome would lead to intracellular n-alkane accumulation, the *Prochlorococcus marinus* MIT9312 adm-aar operon (encoding PMT9312_0532 plus PMT9312_0533) described in Example 7 was chromosomally integrated at the SYNPC7002_A0358 locus. To do so, a SYNPC7002_A0358-targeting vector (pJB1279; SEQ ID NO: 23) was constructed containing 750 bp regions of upstream and downstream homology designed to recombinationally replace the SYNPC7002_A0358 gene with a spectinomycin-resistance cassette downstream of a multiple cloning site (MCS) situated between said regions of homology. Instead of using a constitutive promoter to express the adm-aar operon, an inducible promoter was employed. Specifically, a urea-repressible, nitrate-inducible nirA-type promoter, P(nir07) (SEQ ID NO:24), was inserted into the MCS via NotI and NdeI, generating the base homologous recombination vector pJB1279.

Two operons were cloned downstream of P(nir07) of pJB1279 to generate two experimental constructs, wherein said operons were placed under transcriptional control of P(nir07). The first operon comprised only the aforementioned *Prochlorococcus* PMT9312_0532-PMT9312_0533 operon, inserted via NdeI and EcoRI, resulting in the final plasmid pJB286alk_p; the sequence of this adm-aar operon was exactly as described in Example 7. The second operon comprised (1) the same *Prochlorococcus* PMT9312_0532-PMT9312_0533 adm-aar operon, followed by (2) an adm-aar operon derived from *Cyanothece* sp. ATCC51142 genes cce_0778 (SEQ ID NO: 31) and cce_1430 (SEQ ID NO: 30), respectively, inserted via EcoRI (selecting the correct

orientation by screening), resulting in the final plasmid pJB1256. It is to be noted that *Cyanothece* sp. ATCC51142 naturally synthesizes n-pentadecane as the major intracellular n-alkane. This *Cyanothece* adm-aar operon (SEQ ID NO: 25) was codon- and restriction-site-optimized prior to synthesis by DNA2.0 (Menlo Park, Calif.). The operon expresses proteins with amino acid sequences identical to those of the AAR and ADM enzymes from *Cyanothece* sp. ATCC51142 (SEQ ID NOs: 27 and 29, respectively). The complete operon in plasmid pJB1256, therefore, comprises 4 genes—ADM and AAR from *Prochlorococcus* PMT9312 and ADM and AAR from *Cyanothece* sp. ATCC51142—under the control of a single P(nir07) promoter.

pJB1279, pJB286alk_p, and pJB1256 were naturally transformed into JCC138 exactly as described in Example 5, generating spectinomycin-resistant strains JCC1683c, JCC1683, and JCC1685, respectively. As a first test, a clonal starter culture of each of these three strains, as well as of JCC138, was grown up for 5 days at 37° C. at 150 rpm in 2% CO₂-enriched air at ~100 μmol photons m⁻² s⁻¹ in a Multitron II (Infors) shaking photoincubator in A+ (JCC138) or A+ supplemented with 100 μg/ml spectinomycin (JCC1683c, JCC1683, and JCC1685). At this point, each starter culture was used to inoculate a 30 ml JB2.1 medium plus 3 mM urea flask culture supplemented with no antibiotics (JCC138) or 100 μg/ml spectinomycin (JCC1683c, JCC1683, and JCC1685). The four cultures were then grown for 14 days at 37° C. at 150 rpm in 2% CO₂-enriched air at ~100 μmol photons m⁻² s⁻¹ in a Multitron II (Infors) shaking photoincubator.

20 OD₇₃₀-ml worth of cells was collected by centrifugation in a pre-weighed eppendorf tube. Cells were washed by two cycles of resuspension in Milli-Q water and microcentrifugation, and dewetted by two cycles of microcentrifugation and aspiration. Wet cell pellets were frozen at -80° C. for two hours and then lyophilized overnight, at which point the tube containing the dry cell mass was weighed again such that the mass of the cell pellet (~6 mg) could be calculated within ±0.1 mg. In parallel, 3.5 OD₇₃₀-ml worth of cells from each culture was collected by centrifugation in an eppendorf tube, washed thoroughly by three cycles of resuspension in Milli-Q water and microcentrifugation, and then dewetted as much as possible by three cycles of microcentrifugation and aspiration. Dewetted cell pellets were then extracted by vortexing for 15 seconds in 1.0 ml acetone containing 18.2 mg/l BHT and 16.3 mg/l n-heptacosane (C₂₇) internal standard (ABH); cell debris was pelleted by centrifugation, and 500 μl supernatant was submitted for GC-FID. Acetone-extracted DCW was calculated as 3.5/20, or 17.5%, of the DCW measured for 20 OD₇₃₀-ml worth of cells. In parallel with the four biological

sample extractions, eight empty eppendorf tubes were extracted with ABH in the same fashion. The extraction/injection efficiency of all ABH extractants was assessed by calculating the ratio between the n-heptacosane GC-FID peak area of the sample and the average n-heptacosane GC-FID peak area of the six empty-tube controls—only ratios of 100%±11% were accepted (Table 8).

Concentrations of n-tridecane (C₁₃), n-tetradecane (C₁₄), n-pentadecane (C₁₅), n-hexadecane (C₁₆), n-heptadecane (C₁₇), and n-octadecane (C₁₈), in the four extractants were quantitated by (GC/FID) as described in Example 11. GC-FID conditions were as follows. An Agilent 7890A GC/FID equipped with a 7683 series autosampler was used. 1 μl of each sample was injected into the GC inlet (split 5:1, pressure) and an inlet temperature of 290° C. The column was a HP-5 (Agilent, 30 m×0.32 mm×0.25 μm) and the carrier gas was helium at a flow of 1.0 ml/min. The GC oven temperature program was 50° C., hold 1.0 minute; 10° C./min increase to 290° C.; hold 9 minutes. n-Alkane production was calculated as a percentage of the acetone-extracted DCW. The coefficient of variation of the n-heptacosane GC-FID peak area of the eight empty-tube controls was 3.6%.

GC-FID data are summarized in Table 8. As expected, controls strains JCC138 and JCC1683c made no n-alkanes, whereas JCC683 and JCC1685 made n-alkanes, ~97% of which comprised n-pentadecane and n-heptadecane. JCC1685 made significantly more n-alkanes as a percentage of DCW (~0.42%) compared to JCC1683 (~0.16%), likely explaining the relatively low final OD₇₃₀ of the JCC1685 culture. For JCC1683 expressing *Prochlorococcus marinus* MIT 9312 adm-aar, the percentage by mass of n-pentadecane relative to n-pentadecane plus n-heptadecane was 53.2%, in quantitative agreement with that of JCC1281 expressing the same operon on pAQ1 (57.3%; Table 7). In contrast, for JCC1685 which additionally expresses *Cyanothece* sp. ATCC51142 adm-aar, the percentage by mass of n-pentadecane relative to n-pentadecane plus n-heptadecane was 83.7%. This result demonstrates that the in vivo expression of cce_0778 and cce_1430 in a cyanobacterial host biases the n-alkane product length distribution towards n-pentadecane—even more so than does expression of PMT9312_0532 and PMT9312_0533. The total amount of intracellular n-alkane produced by chromosomal integrants JCC1683 and JCC1685 is apparently lower than that of pAQ1-based transformants such as JCC1469, presumably owing to a combination of lower-copy expression (i.e., chromosome versus high-copy pAQ1), and partially repressed transcription—due to the initial presence of urea in the growth medium—of P(nir07) compared to the constitutive promoters P(aphII) (JCC1281) and P(cI) (JCC1469).

TABLE 8

Strain	OD ₇₃₀	C ₂₇ - normalized extraction/ injection efficiency	C ₁₅ as % of DCW (95% CI%)	C ₁₇ as % of DCW (95% CI%)	(C ₁₅ + C ₁₇)/(C ₁₃ + C ₁₄ + C ₁₅ + C ₁₆ + C ₁₇) Mass %	C ₁₅ / (C ₁₅ + C ₁₇) Mass %
JCC138	17.0	110%	nd	nd	na	na
JCC1683c	13.4	108%	nd	nd	na	na
JCC1683	12.2	111%	0.083% (7.6%)	0.073% (12.5%)	97.3%	53.2%
JCC1685	10.0	110%	0.341% (13.0%)	0.066% (8.8%)	96.7%	83.7%

n-Pentadecane and n-heptadecane quantitated by GC-FID in acetone cell pellet extractants of JCC138, JCC1683c, JCC1683, and JCC1685.

n-Octadecane was not detected in any of the samples;

nd: not detected,

na: not applicable.

In order to confirm the urea-repressibility/nitrate-inducibility of P(nir07), the intracellular n-alkane product distribution of JCC1685 was determined from cultures grown in either JB2.1 medium, containing only nitrate as the nitrogen source, and JB2.1 supplemented with 6 mM urea, urea being preferentially utilized as nitrogen source relative to nitrate and provided at a concentration such that it became depleted when the culture reached an OD₇₃₀ of ~4. JCC1683c in JB2.1 was run in parallel as a negative control. Accordingly, a clonal starter culture of JCC1683c and JCC1685 was grown up for 5 days at 37° C. at 150 rpm in 2% CO₂-enriched air at ~100 μmol photons m⁻² s⁻¹ in a Multitron II (Infors) shaking photoincubator in A+ supplemented with 100 μg/ml spectinomycin. At this point, each starter culture was used to inoculate duplicate 30 ml JB2.1 medium flask cultures supplemented

with 400 μg/ml spectinomycin; in addition, the JCC1685 starter culture was used to inoculate duplicate 30 ml JB2.1 medium plus 6 mM urea flask cultures supplemented with 400 μg/ml spectinomycin. The six cultures were then grown for 14 days at 37° C. at 150 rpm in 2% CO₂-enriched air at ~100 μmol photons m⁻² s⁻¹ in a Multitron II (Infors) shaking photoincubator. Intracellular n-alkanes as a percentage of DCW were determined exactly as described in Example 11; data are summarized in Table 9. Consistent with the urea repressibility of P(nir07), n-alkanes as a percentage of JCC185 DCW were significantly higher in the absence of urea (~0.59%) compared to in the presence of urea (~0.15%). This likely explained the relatively low final OD₇₃₀ of the no-urea cultures.

TABLE 9

Strain	Medium	OD ₇₃₀	C ₂₇ normalized extraction/injection efficiency	C ₁₅ as % of DCW (95% CI%)	C ₁₇ as % of DCW (95% CI%)	n-alkanes as % of DCW	(C ₁₅ + C ₁₇)/(C ₁₃ + C ₁₅ + C ₁₆ + C ₁₇) Mass %	C ₁₅ /C ₁₇ Mass %
JCC1683c #1	JB2.1	9.5	101%	nd	nd	na	na	na
JCC1683c #2	JB2.1	9.5	101%	nd	nd	na	na	na
JCC1685 #1	JB2.1 + 6 mM	7.4	102%	0.076% (7.1%)	0.067% (1.5%)	0.14%	100%	53.2%
JCC1685 #2	JB2.1 + 6 mM	6.4	102%	0.090% (3.3%)	0.051% (2.3%)	0.15%	94.6%	63.9%
JCC1685 #1	JB2.1	1.2	101%	0.42% (1.4%)	0.14% (1.1%)	0.57%	97.9%	74.9%
JCC1685 #2	JB2.1	3.3	102%	0.49% (1.6%)	0.11% (1.6%)	0.60%	100%	81.4%

n-Pentadecane and n-heptadecane quantitated by GC-FID in acetone cell pellet extractants of JCC1683c and JCC1685 as a function of urea in the growth medium
n-Octadecane was not detected in any of the samples;
nd: not detected,
na: not applicable

A number of embodiments of the invention have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the invention. All publications, patents and other references mentioned herein are hereby incorporated by reference in their entirety.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 128

<210> SEQ ID NO 1

<211> LENGTH: 1029

<212> TYPE: DNA

<213> ORGANISM: Thermosynechococcus elongatus

<400> SEQUENCE: 1

```

atgtttggat taattggtea tctgacgagt ctggagcaag cccaagccgt tgcccateag      60
ttgggttacc ccgaatatgc cgatcaaggc ttggaatttt ggtgatgag accgccgcag      120
atcgtcgatg agattacggt gacgagcgta acgggcaaaa ctatctatgg caaatacgtt      180
gagtctctgt ttttaccaga gatgctggcc aaccagcggg tgaaggcagc gactcgcaaa      240
gttattaacg ccatggccca tgcccaaaaag cacaacattg acattacggc cttggggggc      300
ttctcctcga tcattcttga gaactttgat ctggagaaaa tgtcccacat togcaacatt      360
gaactggact ttogccgctt tacaacgggg aataccataa ccgcctatat catctgcaaa      420
caaattgagc aggcggcgcc ccaagtgggg attgatttgc ggcaggcaac cgtggctgtt      480

```

-continued

```

tgtggggcta cgggggatat tggtagtgcc gtctgcccgtt ggttgaatac ctgtttagat 540
gtgcaagatc tcttactcgt agcacggaat cgcgatcgcc tgctggagct acaggcggaa 600
ttgggacggg ggaaaatcct cgacttgatg gaggcgctgc cccttgccga tattgtggtt 660
tgggtggcca gtatgcccac gggagttgag ctgagcattg agcagttaaa acgcccctcc 720
ctgatgattg atgggtggtta tccccaaaat atggccacca aaattcagca cccccagatt 780
catgttctca atgggtgcat tgtcgagcat gccctcgaca ttgactggaa aattatggaa 840
attgtgaata tggatgtgcc ctgcgcccag atgtttgcct gttttgcaga ggctatgctt 900
ttagagttcg agggctggca caccaatttc tcttggggac gcaatcaaat cactgtggaa 960
aagatgcagc aaattggtga ggtctcccgt aaacatggat ttcagccact actgttgaat 1020
cctcagtaa 1029

```

<210> SEQ ID NO 2

<211> LENGTH: 342

<212> TYPE: PRT

<213> ORGANISM: Thermosynechococcus elongatus

<400> SEQUENCE: 2

```

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu His Ala Gln Ala
 1             5             10             15
Val Ala His Gln Leu Gly Tyr Pro Glu Tyr Ala Asp Gln Gly Leu Glu
                20             25             30
Phe Trp Cys Met Ala Pro Pro Gln Ile Val Asp Glu Ile Thr Val Thr
                35             40             45
Ser Val Thr Gly Lys Thr Ile Tyr Gly Lys Tyr Val Glu Ser Cys Phe
                50             55             60
Leu Pro Glu Met Leu Ala Asn Gln Arg Val Lys Ala Ala Thr Arg Lys
 65             70             75             80
Val Ile Asn Ala Met Ala His Ala Gln Lys His Asn Ile Asp Ile Thr
                85             90             95
Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Asn Phe Asp Leu Glu
                100            105            110
Lys Met Ser His Ile Arg Asn Ile Glu Leu Asp Phe Arg Arg Phe Thr
                115            120            125
Thr Gly Asn Thr His Thr Ala Tyr Ile Ile Cys Gln Gln Ile Glu Gln
                130            135            140
Ala Ala Pro Gln Val Gly Ile Asp Leu Arg Gln Ala Thr Val Ala Val
 145            150            155            160
Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Asn
                165            170            175
Thr Cys Leu Asp Val Gln Asp Leu Leu Leu Val Ala Arg Asn Arg Asp
                180            185            190
Arg Leu Leu Glu Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Leu Asp
                195            200            205
Leu Met Glu Ala Leu Pro Leu Ala Asp Ile Val Val Trp Val Ala Ser
                210            215            220
Met Pro Lys Gly Val Glu Leu Ser Ile Glu Gln Leu Lys Arg Pro Ser
 225            230            235            240
Leu Met Ile Asp Gly Gly Tyr Pro Lys Asn Met Ala Thr Lys Ile Gln
                245            250            255
His Pro Gln Ile His Val Leu Asn Gly Gly Ile Val Glu His Ala Leu
                260            265            270

```

-continued

Asp Ile Asp Trp Lys Ile Met Glu Ile Val Asn Met Asp Val Pro Ser
 275 280 285

Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu
 290 295 300

Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Val Glu
 305 310 315 320

Lys Met Gln Gln Ile Gly Glu Val Ser Arg Lys His Gly Phe Gln Pro
 325 330 335

Leu Leu Leu Asn Pro Gln
 340

<210> SEQ ID NO 3
 <211> LENGTH: 696
 <212> TYPE: DNA
 <213> ORGANISM: Thermosynechococcus elongatus

<400> SEQUENCE: 3

```

atgacaacgg ctaccgctac acctgttttg gactaccata gcgatcgcta caaggatgcc      60
tacagccgca ttaacgccat tgtcattgaa ggtgaacagg aagctcacga taactatata      120
gatttagcca agctgctgcc acaacaccaa gaggaactca cccgccttgc caagatggaa      180
gctcgccaca aaaagggggt tgaggcctgt ggtcgcaacc tgagcgtaac gccagatatg      240
gaatttgcca aagccttctt tgaaaaactg cgcgctaact ttcagagggc tctggcggag      300
ggaaaaactg cgacttgtct tctgattcaa gctttgatca tcgaatcctt tgcgatcgcg      360
gctacaaca tctacatccc aatggcggat cctttogccc gtaaaattac tgagagtgtt      420
gttaaggagc aatacagcca cctcaacttt ggcgaaatct ggctcaagga acactttgaa      480
agcgtcaaag gagagctoga agaagccaat cgcgccaatt tacccttggg ctggaaaatg      540
ctcaaccaag tggaagcaga tgccaaagtg ctcggcatgg aaaaagatgc ccttgtggaa      600
gacttcatga ttcagtacag tggtgcccta gaaaatatcg gctttaccac ccgcgaaatt      660
atgaagatgt cagtttatgg cctcactggg gcataa                                696
  
```

<210> SEQ ID NO 4
 <211> LENGTH: 231
 <212> TYPE: PRT
 <213> ORGANISM: Thermosynechococcus elongatus

<400> SEQUENCE: 4

Met Thr Thr Ala Thr Ala Thr Pro Val Leu Asp Tyr His Ser Asp Arg
 1 5 10 15

Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu
 20 25 30

Gln Glu Ala His Asp Asn Tyr Ile Asp Leu Ala Lys Leu Leu Pro Gln
 35 40 45

His Gln Glu Glu Leu Thr Arg Leu Ala Lys Met Glu Ala Arg His Lys
 50 55 60

Lys Gly Phe Glu Ala Cys Gly Arg Asn Leu Ser Val Thr Pro Asp Met
 65 70 75 80

Glu Phe Ala Lys Ala Phe Phe Glu Lys Leu Arg Ala Asn Phe Gln Arg
 85 90 95

Ala Leu Ala Glu Gly Lys Thr Ala Thr Cys Leu Leu Ile Gln Ala Leu
 100 105 110

Ile Ile Glu Ser Phe Ala Ile Ala Ala Tyr Asn Ile Tyr Ile Pro Met
 115 120 125

-continued

Ala Asp Pro Phe Ala Arg Lys Ile Thr Glu Ser Val Val Lys Asp Glu
 130 135 140

Tyr Ser His Leu Asn Phe Gly Glu Ile Trp Leu Lys Glu His Phe Glu
 145 150 155 160

Ser Val Lys Gly Glu Leu Glu Glu Ala Asn Arg Ala Asn Leu Pro Leu
 165 170 175

Val Trp Lys Met Leu Asn Gln Val Glu Ala Asp Ala Lys Val Leu Gly
 180 185 190

Met Glu Lys Asp Ala Leu Val Glu Asp Phe Met Ile Gln Tyr Ser Gly
 195 200 205

Ala Leu Glu Asn Ile Gly Phe Thr Thr Arg Glu Ile Met Lys Met Ser
 210 215 220

Val Tyr Gly Leu Thr Gly Ala
 225 230

<210> SEQ ID NO 5
 <211> LENGTH: 1026
 <212> TYPE: DNA
 <213> ORGANISM: *Synechococcus elongatus*

<400> SEQUENCE: 5

```

atgttcggtc ttatcgggtca tctcaccagt ttggagcagg cccgcgacgt ttctcgcagg      60
atgggctacg acgaatacgc cgatcaagga ttggagtttt ggagtagcgc tectcctcaa      120
atcgttgatg aaatcacagt caccagtgcc acaggcaagg tgattcacgg tcgctacatc      180
gaatcgtggt tcttgccgga aatgctggcg gcgcgccgct tcaaaacagc cacgcgcaaa      240
gttctcaatg ccatgtocca tgccccaaaa caggcatcg acatctcggc cttggggggc      300
tttacctcga ttattttoga gaatttogat ttggccagtt tgcggcaagt gcgcgacact      360
accttgaggt ttgaacggtt caccaccggc aatactcaca cggcctacgt aatctgtaga      420
caggtggaag ccgctgctaa aacgctgggc atcgacatta cccaagcgac agtagcggtt      480
gtcggcgcga ctggcgatat cggtagcgct gtctgcccgt ggctcgacct caaactgggt      540
gtcggtgatt tgatcctgac ggcgcgcaat caggagcgtt tggataacct gcaggctgaa      600
ctcggccggg gcaagattct gcccttgaa gccgctctgc cggaagctga ctttatcgtg      660
tgggtcgcca gtatgcctca gggcgtagt atcgaccag caacctgaa gcaaccctgc      720
gtcctaatcg acgggggcta cccccaaaac ttgggcagca aagtccaagg tgaggcgcac      780
tatgtcctca atggcggggt agttgaacat tgcttcgaca tcgactggca gatcatgtcc      840
gctgcagaga tggcgcggcc cgagcgcag atgtttgcct gctttgccga ggcgatgctc      900
ttggaatttg aaggctggca tactaacttc tcctggggcc gcaaccaaat cacgatcgag      960
aagatggaag cgatcgggtg ggcacgggtg cgccacggct tccaaccctt ggcattggca    1020
atttga                                             1026

```

<210> SEQ ID NO 6
 <211> LENGTH: 341
 <212> TYPE: PRT
 <213> ORGANISM: *Synechococcus elongatus*

<400> SEQUENCE: 6

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu Gln Ala Arg Asp
 1 5 10 15

Val Ser Arg Arg Met Gly Tyr Asp Glu Tyr Ala Asp Gln Gly Leu Glu
 20 25 30

-continued

Phe Trp Ser Ser Ala Pro Pro Gln Ile Val Asp Glu Ile Thr Val Thr
 35 40 45

Ser Ala Thr Gly Lys Val Ile His Gly Arg Tyr Ile Glu Ser Cys Phe
 50 55 60

Leu Pro Glu Met Leu Ala Ala Arg Arg Phe Lys Thr Ala Thr Arg Lys
 65 70 75 80

Val Leu Asn Ala Met Ser His Ala Gln Lys His Gly Ile Asp Ile Ser
 85 90 95

Ala Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asp Leu Ala
 100 105 110

Ser Leu Arg Gln Val Arg Asp Thr Thr Leu Glu Phe Glu Arg Phe Thr
 115 120 125

Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu Ala
 130 135 140

Ala Ala Lys Thr Leu Gly Ile Asp Ile Thr Gln Ala Thr Val Ala Val
 145 150 155 160

Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Asp
 165 170 175

Leu Lys Leu Gly Val Gly Asp Leu Ile Leu Thr Ala Arg Asn Gln Glu
 180 185 190

Arg Leu Asp Asn Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Leu Pro
 195 200 205

Leu Glu Ala Ala Leu Pro Glu Ala Asp Phe Ile Val Trp Val Ala Ser
 210 215 220

Met Pro Gln Gly Val Val Ile Asp Pro Ala Thr Leu Lys Gln Pro Cys
 225 230 235 240

Val Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Ser Lys Val Gln
 245 250 255

Gly Glu Gly Ile Tyr Val Leu Asn Gly Gly Val Val Glu His Cys Phe
 260 265 270

Asp Ile Asp Trp Gln Ile Met Ser Ala Ala Glu Met Ala Arg Pro Glu
 275 280 285

Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu
 290 295 300

Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Ile Glu
 305 310 315 320

Lys Met Glu Ala Ile Gly Glu Ala Ser Val Arg His Gly Phe Gln Pro
 325 330 335

Leu Ala Leu Ala Ile
 340

<210> SEQ ID NO 7
 <211> LENGTH: 696
 <212> TYPE: DNA
 <213> ORGANISM: Synechococcus elongatus

<400> SEQUENCE: 7

atgccgcagc ttgaagccag ccttgaactg gactttcaaa gcgagtccta caaagacgct	60
tacagccgca tcaacgcgat cgtgattgaa ggcaacaag aggcgttcga caactacaat	120
cgccctgctg agatgctgcc cgaccagcgg gatgagcttc acaagctagc caagatggaa	180
cagcgccaca tgaaggcgtt tatggcctgt ggcaaaaatc tctccgtcac tctgacatg	240
ggttttgccc agaaatttt cgagcgcttg cacgagaact tcaaagcggc ggctgcggaa	300
ggcaaggtcg tcacctgctt actgattcaa tcgctaatac tcgagtgcct tgcgatgcg	360

-continued

```

gcttacaaca tctacatccc agtggcggat gcttttgccc gcaaaatcac ggagggggtc 420
gtgcgcgaag aatacctgca ccgcaacttc ggtgaagagt ggctgaaggc gaattttgat 480
gcttccaaag ccgaactgga agaagccaat cgtcagaacc tgccttggt ttggctaagt 540
ctcaacgaag tggccgatga tgctcgcaa ctcgggatgg agegtgagtc gctcgtcgag 600
gactttatga ttgcctacgg tgaagctctg gaaaacatcg gcttcacaac gcgcgaaatc 660
atgcgtatgt ccgcctatgg ccttgcgccc gtttga 696

```

```

<210> SEQ ID NO 8
<211> LENGTH: 231
<212> TYPE: PRT
<213> ORGANISM: Synechococcus elongatus

```

```

<400> SEQUENCE: 8

```

```

Met Pro Gln Leu Glu Ala Ser Leu Glu Leu Asp Phe Gln Ser Glu Ser
1          5          10          15
Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu
20          25          30
Gln Glu Ala Phe Asp Asn Tyr Asn Arg Leu Ala Glu Met Leu Pro Asp
35          40          45
Gln Arg Asp Glu Leu His Lys Leu Ala Lys Met Glu Gln Arg His Met
50          55          60
Lys Gly Phe Met Ala Cys Gly Lys Asn Leu Ser Val Thr Pro Asp Met
65          70          75          80
Gly Phe Ala Gln Lys Phe Phe Glu Arg Leu His Glu Asn Phe Lys Ala
85          90          95
Ala Ala Ala Glu Gly Lys Val Val Thr Cys Leu Leu Ile Gln Ser Leu
100         105         110
Ile Ile Glu Cys Phe Ala Ile Ala Ala Tyr Asn Ile Tyr Ile Pro Val
115         120         125
Ala Asp Ala Phe Ala Arg Lys Ile Thr Glu Gly Val Val Arg Asp Glu
130         135         140
Tyr Leu His Arg Asn Phe Gly Glu Glu Trp Leu Lys Ala Asn Phe Asp
145         150         155         160
Ala Ser Lys Ala Glu Leu Glu Glu Ala Asn Arg Gln Asn Leu Pro Leu
165         170         175
Val Trp Leu Met Leu Asn Glu Val Ala Asp Asp Ala Arg Glu Leu Gly
180         185         190
Met Glu Arg Glu Ser Leu Val Glu Asp Phe Met Ile Ala Tyr Gly Glu
195         200         205
Ala Leu Glu Asn Ile Gly Phe Thr Thr Arg Glu Ile Met Arg Met Ser
210         215         220
Ala Tyr Gly Leu Ala Ala Val
225         230

```

```

<210> SEQ ID NO 9
<211> LENGTH: 1041
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        polynucleotide

```

```

<400> SEQUENCE: 9

```

```

atgtttggtc tgattggcca tagcaccagc tttgaggacg caaagcgcaa ggcgagcctg 60

```

-continued

```

ctgggtttcg accacatcgc ggatggcgat ctggatgtgt ggtgtaccgc accgccgcaa 120
ctggttgaaa acgtggaagt caaaagcgcg acgggtatca gcattgaagg tagctatata 180
gatagctgct tcgtgccgga gatgctgagc cgcttcaaga ccgcgcgctcg taaagttctg 240
aatgcaatgg agctggcgca gaaaaagggat atcaatatca ctgcccctggg tggetttacc 300
tccattatct ttgagaactt caacctggtg cagcacaagc aaatccgtaa taccagcctg 360
gagtgggagc gtttcaccac gggtaacacg cacacggcat gggtgatttg tcgtcagctg 420
gagatcaacg caccgcgcat tggcatcgac ctgaaaactg caacggtcgc tgttatcggc 480
cgcaccggcg atattgtag cgcggtgtgt cgctggctgg tcaataagac cggcattagc 540
gaactgctga tggctgctcg ccaacaacag ccaactgacct tgctgcaaaa agaactggac 600
ggtggcacca tcaagagcct ggatgaagcc ctgccgcagg cggatattgt cgtgtgggtt 660
gcttcgatgc ctaagacgat cgaattgag attgaaaacc tgaaaaagcc gtgcctgatg 720
atcgacggtg gctaccgcaa gaactcggac gagaaattca aaggcaaaaa cattcacgtg 780
ttgaagggtg gtatcgtcga gtttttcaac gacattggct ggaacatgat ggagttggcg 840
gagatgcaaa acccgcagcg tgagatgttt cegtgcctcg ccgaagctat gattctggag 900
tttgagaaat gccataccaa ctttagctgg ggccgtaaca atatcagctt ggagaagatg 960
gagttcatcg gtgctgcac tctgaagcac ggtttcagcg cgatcggctt ggataaacag 1020
ccgaaagtct tgaccgtttg a 1041

```

<210> SEQ ID NO 10

<211> LENGTH: 345

<212> TYPE: PRT

<213> ORGANISM: *Prochlorococcus marinus*

<400> SEQUENCE: 10

```

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys Arg
1           5           10          15
Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu Asp
20          25          30
Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val Lys
35          40          45
Ser Ala Thr Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys Phe
50          55          60
Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
65          70          75          80
Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala Leu
85          90          95
Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln His
100         105         110
Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr Gly
115         120         125
Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Ile Asn Ala
130         135         140
Pro Arg Ile Gly Ile Asp Leu Lys Thr Ala Thr Val Ala Val Ile Gly
145         150         155         160
Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Val Asn Lys
165         170         175
Thr Gly Ile Ser Glu Leu Leu Met Val Ala Arg Gln Gln Gln Pro Leu
180         185         190
Thr Leu Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Ser Leu Asp

```

-continued

195					200					205					
Glu	Ala	Leu	Pro	Gln	Ala	Asp	Ile	Val	Val	Trp	Val	Ala	Ser	Met	Pro
210					215					220					
Lys	Thr	Ile	Glu	Ile	Glu	Ile	Glu	Asn	Leu	Lys	Lys	Pro	Cys	Leu	Met
225					230					235					240
Ile	Asp	Gly	Gly	Tyr	Pro	Lys	Asn	Leu	Asp	Glu	Lys	Phe	Lys	Gly	Lys
				245					250					255	
Asn	Ile	His	Val	Leu	Lys	Gly	Gly	Ile	Val	Glu	Phe	Phe	Asn	Asp	Ile
			260					265						270	
Gly	Trp	Asn	Met	Met	Glu	Leu	Ala	Glu	Met	Gln	Asn	Pro	Gln	Arg	Glu
			275					280						285	
Met	Phe	Ala	Cys	Phe	Ala	Glu	Ala	Met	Ile	Leu	Glu	Glu	Lys	Cys	His
290					295					300					
Thr	Asn	Phe	Ser	Trp	Gly	Arg	Asn	Asn	Ile	Ser	Leu	Glu	Lys	Met	Glu
305					310					315					320
Phe	Ile	Gly	Ala	Ala	Ser	Leu	Lys	His	Gly	Phe	Ser	Ala	Ile	Gly	Leu
				325					330					335	
Asp	Lys	Gln	Pro	Lys	Val	Leu	Thr	Val							
			340					345							

<210> SEQ ID NO 11
 <211> LENGTH: 756
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 11

```

atgcacaatg aattgaaat cacggatg caaacgctgg aaaccaaac caagacgacc   60
gaagagtcta ttgacaccaa tagcctgaac ctgccggact ttactaccga cagctacaag   120
gatgcctatt ctgcattaa cgccatcgtt attgagggcg aacaggaagc tcatgacaat   180
tacatctcca tcgcaacgct gatcccgat gagctggaag agctgacgaa gctggcacgt   240
atggagctga aacacaagaa aggttttact gcgtgcggtc gtaatctggg tgtggacgca   300
gacatggttt tcgcgaaaaa gttcttcagc aaactgcacg gcaatttcca aatcgcgctg   360
gaaaaaggtg acctgaccac ctgcttgctg atccaagcga ttctgatcga agcatttgcg   420
atttccgcgt acaatgttta catccgtgtg gccgaccat ttgccccaaa gattaccgag   480
gggtgtgtca aagacgagta tctgcatctg aactatggtc aggagtggct gaaaaagaat   540
ctgtccacgt gtaaagaaga gctgatggag gccacaagg tcaatctgcc gctgattaag   600
aaaatgctgg acgaagtggc agaagatgag agcgttttgg cgatggatcg tgaagagttg   660
atggaagagt tcatgattgc gtaccaggat accctgttgg agattggcct ggataatcgc   720
gaaattgccc gtatggcgat ggcggccatt gttag                                     756

```

<210> SEQ ID NO 12
 <211> LENGTH: 251
 <212> TYPE: PRT
 <213> ORGANISM: Prochlorococcus marinus

<400> SEQUENCE: 12

Met	His	Asn	Glu	Leu	Lys	Ile	Thr	Asp	Met	Gln	Thr	Leu	Glu	Thr	Asn
1				5					10					15	
Thr	Lys	Thr	Thr	Glu	Glu	Ser	Ile	Asp	Thr	Asn	Ser	Leu	Asn	Leu	Pro
				20				25						30	

-continued

Asp Phe Thr Thr Asp Ser Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala
 35 40 45
 Ile Val Ile Glu Gly Glu Gln Glu Ala His Asp Asn Tyr Ile Ser Ile
 50 55 60
 Ala Thr Leu Ile Pro Asn Glu Leu Glu Glu Leu Thr Lys Leu Ala Arg
 65 70 75 80
 Met Glu Leu Lys His Lys Lys Gly Phe Thr Ala Cys Gly Arg Asn Leu
 85 90 95
 Gly Val Asp Ala Asp Met Val Phe Ala Lys Lys Phe Phe Ser Lys Leu
 100 105 110
 His Gly Asn Phe Gln Ile Ala Leu Glu Lys Gly Asn Leu Thr Thr Cys
 115 120 125
 Leu Leu Ile Gln Ala Ile Leu Ile Glu Ala Phe Ala Ile Ser Ala Tyr
 130 135 140
 Asn Val Tyr Ile Arg Val Ala Asp Pro Phe Ala Lys Lys Ile Thr Glu
 145 150 155 160
 Gly Val Val Lys Asp Glu Tyr Leu His Leu Asn Tyr Gly Gln Glu Trp
 165 170 175
 Leu Lys Lys Asn Leu Ser Thr Cys Lys Glu Glu Leu Met Glu Ala Asn
 180 185 190
 Lys Val Asn Leu Pro Leu Ile Lys Lys Met Leu Asp Glu Val Ala Glu
 195 200 205
 Asp Ala Ser Val Leu Ala Met Asp Arg Glu Glu Leu Met Glu Glu Phe
 210 215 220
 Met Ile Ala Tyr Gln Asp Thr Leu Leu Glu Ile Gly Leu Asp Asn Arg
 225 230 235 240
 Glu Ile Ala Arg Met Ala Met Ala Ala Ile Val
 245 250

<210> SEQ ID NO 13

<211> LENGTH: 1041

<212> TYPE: DNA

<213> ORGANISM: Prochlorococcus marinus

<400> SEQUENCE: 13

atgtttgggt taataggcca ctcaactagt tttgaagatg caaaaagaaa agcttcatta 60
 ctaggctttg atcatattgc tgatggtgat ctgatgttt ggtgtacagc ccctcctcaa 120
 ttggttgaaa atgtagaagt taagagtgct actggaatat ctattgaagg ttcttatata 180
 gattcttget ttgttcctga aatgctttct aggtttaaaa ccgcaagaag aaaagtatta 240
 aatgctatgg aattagctca gaaaaaaggg attaacatta cggctttagg aggatttact 300
 tctattattt tcgaaaattt taatcttctt caacataaac aaattagaaa tacttcatta 360
 gagtgggaaa ggtttactac aggtaataca cacactgcct gggttatttg taggcaacta 420
 gaaataaatg ctccctcgcat agggatagat cttaaaaactg caactgttgc tgttattggt 480
 gctacaggtg atataggaag tgetgtttgt aggtggcttg tcaataaaac tggattttca 540
 gaacttctta tgggtgctag acaacaacaa ccattaactc tattacagaa agaattagat 600
 ggtggcacta taaaaagttt agatgaagca ttgcctcaag cggatattgt tgtatgggtt 660
 gcaagtatgc ctaaaacgat tgaattgaa attgaaaact taaaaaaccc atgtttaatg 720
 attgatggtg gataccctaa aaatcttgat gagaaattta aaggtaaaaa tattcatggt 780
 ttaaaaggag gtatagtaga gtttttcaat gatattggct ggaatatgat ggaacttgca 840

-continued

```

gaaatgcaga accctcagag agagatgttt gcttgctttg cagaagctat gatttttagaa 900
tttgaaaagt gtcataccaa ctttagttgg ggaaggaata acattttctct tgaaaaaatg 960
gaattttatg gagcagcttc tttgaaacat ggtttttctg cgattggact tgataaacag 1020
cctaaagtat tgactgtttg a 1041

```

```

<210> SEQ ID NO 14
<211> LENGTH: 756
<212> TYPE: DNA
<213> ORGANISM: Prochlorococcus marinus

```

```

<400> SEQUENCE: 14

```

```

atgcataatg agctaaagat tactgacatg caaactctag aaacaaatac aaaaactact 60
gaagaatcca tagacacgaa ttctttgaat cttcccgact ttacaacaga ttcctataag 120
gatgcatata gcagaataaa tgcaattggt atagagggag agcaagaggc tcatgataat 180
tacatttcaa tagcaacggt aataccaaat gagttagaag aattaactaa gttggcgaga 240
atggaactca agcataaaaa aggatttact gcttgtggaa gaaatttagg agtagatgct 300
gatatggtat tcgcaaaaaa attcctttct aaattgcatg gtaattttca aattgcttta 360
gaaaaaggaa atttaacaac ttgtctctg atacaagcta ttttaattga agcttttget 420
atatctgctt ataacgttta cataagagtt gctgatcctt ttgcaaaaaa aataacagag 480
ggagtgggta aagatgaata tctccatcta aattacggcc aagagtggct taaaaagaat 540
ttatctactt gtaagaaga attaatggaa gccataaagg ttaaccttc cttaattaaa 600
aagatgtag atgaagtagc agaagatgca tcagttttgg ctatggatag agaagagtta 660
atggaagaat ttatgattgc ttaccaagac actcttctag aaataggctc tgataataga 720
gaaattgcaa gaatggctat ggcagcgatt gtttaa 756

```

```

<210> SEQ ID NO 15
<211> LENGTH: 7026
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

```

```

<400> SEQUENCE: 15

```

```

gtcagcaagc tctggaatth cccgattctc tgatgggaga tccaaaaatt ctgcagctcc 60
ctcaatcacg atatcggctc tggatcgccc tgtagcttcc gacaactgct caattttttc 120
gagcatctct accgggcacg ggaatgaaat taacgggtgt ttagccatgt gttatacagt 180
gtttacaact tgactaacia atacctgcta gtgtatacat attgtattgc aatgtatacg 240
ctattttccg tgetgtcttt aatgggggatt atcgcaagca agtaaaaaag cctgaaaacc 300
ccaataggta agggattccg agcttactcg ataattatca cctttgagcg cccctaggag 360
gaggcgaaaa gctatgtctg acaaggggtt tgaccctga agtcggttgcg cgagcattaa 420
ggtctgcgga tagcccataa catacttttg ttgaacttgt gcgcttttat caaccctta 480
agggcttggg agcgttttat gcggccgcgg gggggggggg gaaagccacg ttgtgtctca 540
aaatctctga tgttacattg cacaagataa aaatatatca tcatgaacia taaaactgtc 600
tgcttacata aacagtaata caaggggtca tatgcccag cttgaagcca gccttgaact 660
ggactttcaa agcgagtctc acaagacgc ttacagccgc atcaacgcga tcgtgattga 720
aggcgaacia gaggcgcttcg acaactacia tcgccttctc gagatgtgc ccgaccagcg 780

```

-continued

ggatgagctt cacaagctag ccaagatgga acagcgcac atgaaaggct ttatggcctg	840
tggcaaaaat ctctccgca ctectgacat gggttttgcc cagaaatfff tcgagcgctt	900
gcacgagaac ttcaaacggg cggtcgcca aggcaaggtc gtcacctgcc tactgattca	960
atcgctaadc atcgagtgct ttgcgatcgc ggcttacaac atctacatcc cagtggcgga	1020
tgcttttgcc cgcaaaatca cggagggggg cgtgcgcgac gaataacctgc accgcaactt	1080
cgggtaagag tggctgaagg cgaatfffga tgcttccaaa gccgaactgg aagaagccaa	1140
tcgtcagaac ctgcccttgg tttggctaata gctcaacgaa gtggccgatg atgctcgca	1200
actcgggatg gagcgtgagt cgctcgcga ggactttatg attgcctacg gtgaagctct	1260
ggaaaacatc ggcttcacaa cgcgcgaaat catgcgatg tccgcctatg gccttgcggc	1320
cgtttgatcc aggaaatctg aatgctggg cttatcggtc atctcaccag tttggagcag	1380
gccccgacg tttctcgca gatgggctac gacgaatacg ccgatcaagg attggagttt	1440
tggagtagcg ctctctca aatcgtgat gaaatcacag tcaccagtgc cacaggcaag	1500
gtgattcaag gtcgctacat cgaatcgtgt ttcttgccgg aaatgctggc ggcgcgcgc	1560
ttcaaaacag ccacgcgcaa agttctcaat gccatgtccc atgccaaaa acacggcatc	1620
gacatctcgg ccttgggggg ctttacctcg attatfffcg agaatffoga tttggccagt	1680
ttgcggcaag tgcgcgacac taccttgag tttgaacggc tcaccaccgg caatactcac	1740
acggcctacg taatctgtag acaggtgga gccgctgcta aaacgctggg catcgacatt	1800
acccaagcga cagtagcggg tgcgcgcgc actggcgata tcggtagcgc tgtctgcgc	1860
tggctcgacc tcaaacggg tgcggtgat ttgatcctga cggcgcgcaa tcaggagcgt	1920
ttggataacc tgcaggctga actcggccgg ggcaagattc tgcccttggg agcgcctctg	1980
cgggaagctg actttatcgt gttggctgcc agtatgcctc agggcgtagt gatcgacca	2040
gcaaccctga agcaaccctg cgtcctaate gacgggggct acccaaaaa cttgggcagc	2100
aaagtccaag gtgagggcat ctatgtctc aatggcggg tagttgaaca ttgcttcgac	2160
atcgactggc agatcatgct cgctgcagag atggcgcggc ccgagcgcca gatgtttgc	2220
tgctttgcgg aggcgatgct cttggaattt gaagcctggc atactaactt ctctggggc	2280
cgcaacccaa tcacgatcga gaagatgga gcgatcggg aggcacggg gcgccacggc	2340
ttccaacctc tggcattggc aatftgaggt ctgtgaattc ggttttccgt cctgtcttga	2400
ttttcaagca aacaatgctt ccgatttcta atcgaggca tttgttttg tttattgcaa	2460
aaacaaaaaa tattgttaca aatftttaca ggctattaag cctaccgtca taaataattt	2520
gccatttact agtttttaat taaccagaac cttgaccgaa cgcagcggg gtaacggcgc	2580
agtggcgggt ttcattgctt gttatgactg tttttttgg gtacagtcta tgccctgggc	2640
atccaagcag caagcgcgtt acgcctggg tcgatgtttg atgttatgga gcagcaacga	2700
tgttacgcag cagggcagtc gccctaaaac aaagttaaac atcatgaggg aagcgggtgat	2760
cgccgaagta tcgactcaac tatcagaggt agttggcgtc atcgagcgc atctcgaacc	2820
gacgttgctg gccgtacatt tgcacggctc cgcagtgat ggcggcctga agccacacag	2880
tgatattgat ttgctgggta cggtgaccgt aaggcttgat gaaacaacgc ggcgagcttt	2940
gatcaacgac cttttggaaa cttcggcttc cctgggagag agcgagattc tccgcctgtg	3000
agaagtcaac attgttgctc acgacgacat cattccgtgg cgttatccag ctaagcgcga	3060
actgcaattt ggagaatggc agcgaatga cattcttga ggtatcttc agccagccac	3120
gatcgacatt gatctggcta tcttgctgac aaaagcaaga gaacatagcg ttgccttggg	3180

-continued

aggtccagcg	gcgagggaac	tctttgatcc	ggttctgaa	caggatctat	ttgaggcgct	3240
aatgaaacc	ttaacgctat	ggaactcgcc	gcccactgg	gctggcgatg	agcgaaatgt	3300
agtgttacg	ttgtcccga	tttggtagc	cgtagtaacc	ggcaaatcg	cgccgaagga	3360
tgctgctgcc	gactgggcaa	tggagcgct	gcccggccag	tatcagccc	tcatacttga	3420
agctagacag	gcttatcttg	gacaagaaga	agatcgcttg	gcctcgcgcg	cagatcagtt	3480
ggaagaatth	gtccactacg	tgaaggcga	gatcaccaag	gtagtcggca	aataatgtct	3540
aacaattcgt	tcaagccgac	gcccgttcgc	ggcgcgctt	aactcaagcg	ttagatgcac	3600
taagcacata	attgctcaca	gccaactat	caggtaagt	ctgcttttat	tatttttaag	3660
cgtgcataat	aagccctaca	caaattggga	gatatatcat	gaggcgcgcc	acgagtgccg	3720
ggaaatttcg	ggggcgatcg	cccctatct	gcaaaaagga	gttaccocat	cagagctata	3780
gtcgagaaga	aaacctcat	tcactcaaca	aggctatgtc	agaagagaaa	ctagaccgga	3840
tcgaagcagc	cctagagcaa	ttggataagg	atgtgcaaac	gctccaaaca	gagcttcagc	3900
aatcccaaaa	atggcaggac	aggacatggg	atgttgtaa	gtgggtaggc	ggaatctcag	3960
cgggcctagc	ggtgagcgct	tccattgcc	tgctcggtt	ggtctttaga	tttctgttt	4020
ccctgccata	aaagcacatt	cttataagtc	atcttgctt	acatcaagga	acaaaaacg	4080
cattgtgctt	tgcaaggcac	aatgtcttc	tcttatgcac	agatggggac	tggaaaccac	4140
acgcacaatt	cccttaaaaa	gcaaccgcaa	aaaataacca	tcaaaataaa	actggacaaa	4200
ttctcatgtg	ggccggccaa	aatgaagtga	agttctata	ctttctagag	aataggaact	4260
tctatagtga	gtcgaataag	ggcgacacaa	aatttattct	aatgcataa	taaatactga	4320
taacatctta	tagtttgtat	tatattttgt	attatcgttg	acatgtataa	tttgatatac	4380
aaaaactgat	ttcccttta	ttattttcga	gatttatttt	cttaattctc	tttaacaaac	4440
tagaaatatt	gtatatacaa	aaaatcataa	ataatagatg	aatagtttaa	ttataggtgt	4500
tcatcaatcg	aaaaagcaac	gtatcttatt	taaagtgcgt	tgcttttttc	tcatttataa	4560
ggttaataaa	ttctcatata	tcaagcaaag	tgacaggcgc	ccttaaatat	tctgacaaat	4620
gctctttccc	taaaactccc	ccataaaaaa	acccgcccga	gcggtttttt	acgttatttg	4680
cggattaacg	attactcgtt	atcagaaccg	cccagggggc	ccgagcttaa	gactggccgt	4740
cgttttacaa	cacagaaaaga	gtttgtagaa	acgcaaaaag	gccatccgct	aggggccttc	4800
tgcttagttt	gatgcctggc	agttccctac	tctgccttc	cgcttctctg	ctcactgact	4860
cgctgcgctc	ggtcgctcgg	ctgcggcgag	cggtatcagc	tactcaaaag	gcgtaatac	4920
ggttatccac	agaatcaggg	gataacgcag	gaaagaacat	gtgagcaaaa	ggccagcaaa	4980
agggcaggaa	ccgtaaaaag	gcccgttgct	tggcgttttt	ccataggctc	cgccccctg	5040
acgagcatca	caaaaatcga	cgctcaagtc	agaggtggcg	aaaccgcaca	ggactataaa	5100
gataccaggc	gtttccccct	ggaagctccc	tctgctgctc	tcctgttccg	accctgccgc	5160
ttaccggata	cctgtccgcc	tttctccctt	cgggaagcgt	ggcgctttct	catagctcac	5220
gctgtaggta	tctcagttcg	gtgtaggtcg	tctcctccaa	gctgggctgt	gtgcacgaac	5280
ccccggttea	gcccgaccgc	tgcccttat	ccgtaacta	tctgtttgag	tccaaccg	5340
taagacacga	cttatcgcca	ctggcagcag	ccactggtaa	caggattagc	agagcgaggt	5400
atgtaggcgg	tgctacagag	ttcttgaagt	ggtgggctaa	ctacggctac	actagaagaa	5460
cagtatttgg	tatctgcgct	ctgctgaagc	cagttacctt	cggaaaaaga	ggtgtagct	5520

-continued

```

cttgatccgg caaaciaaac accgctggta gcggtggttt tttgtttgc aagcagcaga 5580
ttacgcgcag aaaaaaagga tctcaagaag atcctttgat cttttctacg gggctctgacg 5640
ctcagtggaa cgacgcgcgc gtaactcacg ttaaggatt ttggtcatga gcttgccgccg 5700
tcccgtcaag tcagcgtaat gctctgcttt taccaatgct taatcagtga ggcacctatc 5760
tcagcgtatc gtctatttcg ttcacccata gttgcctgac tcccgcctg gttagataact 5820
acgatacggg agggccttacc atctggcccc agcgcctgcga tgataccgcg agaaccacgc 5880
tcaccggctc cggatttata agcaataaac cagccagccg gaagggccga gcgcagaagt 5940
ggtcctgcaa ctttatccgc ctccatccag tctattaatt gttgccggga agctagagta 6000
agtagttcgc cagttaatag tttgcgcaac gttgttgcca tcgctacagg catcgtggtg 6060
tcacgctcgt cgtttggtat ggcttcattc agctccggtt cccaacgac aaggcgagtt 6120
acatgatccc ccatggttggt caaaaaagcg gttagctcct tcggtcctcc gatcgttgtc 6180
agaagtaagt tggccgcagt gttatcactc atggttatgg cagcaactgca taattctctt 6240
actgtcatgc catccgtaag atgcttttct gtgactggtg agtactcaac caagtcattc 6300
tgagaatagt gtatgcggcg accgagttgc tcttgcccgg cgtaaatcag ggataatacc 6360
gcgccacata gcagaacttt aaaagtgtc atcattggaa aacgttcttc gggggcaaaa 6420
ctctcaagga tcttaccgct gttgagatcc agttcagatg aacctactcg tgcacccaac 6480
tgatcttcag catcttttac tttcaccagc gtttctgggt gagcaaaaac aggaaggcaa 6540
aatgccgcaa aaaagggaaat aagggcgaca cggaaatggt gaatactcat attcttctt 6600
tttcaatatt attgaagcat ttatcagggt tattgtctca tgagcggata catattgaa 6660
tgtatttaga aaaataaaca aataggggtc agtgttacaa ccaattaacc aattctgaac 6720
attatcgcga gccattttat acctgaatat ggctcataac accccttgtt tgccctggcg 6780
cagtagcgcg gtggtcccac ctgaccccat gccgaactca gaagtgaac gccgtagcgc 6840
cgatggtagt gtggggactc cccatgcgag agtagggaac tgccaggcat caaataaaac 6900
gaaaggctca gtcgaaagac tgggccttcc gcccgggcta attatgggtt gtcgccccta 6960
ttcgactcta tagtgaagtt cctattctct agaaagtata ggaacttctg aagtggggcc 7020
tgcagg 7026

```

<210> SEQ ID NO 16

<211> LENGTH: 5513

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 16

```

ggggaattgt gagcggataa caattcccct gtagaaataa tttgttttaa cttaataag 60
gagatatacc atgggcgagca gccatcacca tcatcaccac agccaggatc cgaattcgag 120
ctcggcgcgc ctgcaggtcg acaagcttgc ggccgcataa tgcttaagtc gaacagaaag 180
taatcgtatt gtacacggcc gcataatcga aattaatacg actcactata ggggaattgt 240
gagcggataa caattcccga tcttagtata ttagttaagt ataagaagga gatatacata 300
tgccgcagct tgaagccagc cttgaactgg actttcaaag cgagtcctac aaagacgctt 360
acagccgcat caacgcgac gtgattgaag gcgaacaaga ggcgttcgac aactacaatc 420
gccttgctga gatgctgccc gaccagcggg atgagcttca caagctagcc aagatggaac 480

```

-continued

agcgccacat gaaaggcttt atggcctgtg gcaaaaatct ctccgtcact cctgacatgg	540
gttttgccca gaaatttttc gagcgcttgc acgagaactt caaagcggcg gctgcggaag	600
gcaaggctcg cacctgccta ctgattcaat cgctaatacat cgagtgcctt gcgatcgcg	660
cttacaacat ctacatccca gtggcggatg cttttgcccg caaaatcacg gagggggtcg	720
tgcgcgacga atacctgcac cgcaacttcg gtgaagagtg gctgaaggcg aattttgatg	780
cttccaaagc cgaactggaa gaagccaatc gtcagaacct gcccttggtt tggctaatac	840
tcaacgaagt ggccgatgat gctcgcgaac tcgggatgga gcgtgagtcg ctcgtcgagg	900
actttatgat tgcctacggt gaagctctgg aaaacatcgg cttcacaacg cgcgaaatca	960
tgcgtatgtc cgcctatggc cttgcccggc tttgatccag gaaatctgaa tgttcggtct	1020
tatcggatcat ctcaccagtt tggagcaggc ccgcgacgtt tctcgcagga tgggctacga	1080
cgaatacacc gatcaaggat tggagttttg gagtagcgtt cctcctcaaa tcggttatga	1140
aatcacagtc accagtgcc aaggcaaggt gattcacggt cgctacatcg aatcgtggtt	1200
cttgccggaa atgctggcgg cgcgccgctt caaaacagcc acgcgcaaag ttctcaatgc	1260
catgtcccat gcccaaaaac acggcatcga catctcggcc ttggggggct ttacctcgat	1320
tattttcgag aatttcgatt tggccagttt gcggcaagtg cgcgacacta cettggagtt	1380
tgaacggttc accaccggca atactcacac ggccctacgta atctgtagac aggtggaagc	1440
cgctgctaaa acgctgggca tcgacattac ccaagcgaca gtagcggttg tcggcgcgac	1500
tggcgataac ggtagcgtg tctgccgtg gctcgcctc aaactgggtg tcggtgattt	1560
gatcctgacg gcgcgcaatc aggagcgtt ggataacctg caggctgaac tcggccgggg	1620
caagattctg cccttggaag ccgctctgcc ggaagctgac tttatcgtg gggtcgccag	1680
tatgcctcag ggcgtagtga tcgaccacgc aacctgaag caacctcgcg tctaatcga	1740
cgggggctac cccaaaaact tgggcagcaa agtccaaggt gagggcatct atgtcctcaa	1800
tggcgggta gttgaacatt gcttcgacat cgactggcag atcatgtccg ctgcagagat	1860
ggcgcgccc gagcgccaga tgtttgctg ctttgccgag gcgatgctct tggaaattga	1920
aggctggcat actaacttct cctggggcgg caaccaaatc acgatcgaga agatggaagc	1980
gatcggtagc gcacgcgtgc gccacggctt ccaaccttg gcattggcaa tttgaggtct	2040
gtgaattgga tatcggccgg ccacgcgacg gctgacgtcg gtacctcga gtctggtaaa	2100
gaaaccgctg ctgcgaaatt tgaacgccag cacatggact cgtctactag cgcagcttaa	2160
ttaacctagg ctgctgccac cgtgagcaa taactagcat aacctctgg ggcctctaaa	2220
cgggtcctga ggggtttttt gctgaaacct caggcatttg agaagcacac ggtcacactg	2280
cttccgtag tcaataaacc ggtaaaccag caatagacat aagcggctat ttaacgacct	2340
tgcctgaac cgaacgaccg gtcacatcgg ccggatcttg cggcccctcg gcttgaacga	2400
attgtagtac attatttgc gactaccttg gtgatctcgc ctttcacgta gtggacaaat	2460
tcttccaact gatctgcgcg cgaggccaag cgatcttctt cttgtccaag ataagcctgt	2520
ctagctcaa gtatgacggg ctgatactgg gccgcaggc gctccattgc ccagtcggca	2580
gcgacatcct tcggcgcgat tttgccggtt actgcgctgt accaaatgcg ggacaacgta	2640
agcactacat ttcgctcacc gccagcccag tcgggcggcg agttccatag cgttaagggt	2700
tcatttagcg cctcaaatag atcctgttca ggaaccggat caaagagttc ctccgcgct	2760
ggacctacca aggcaacgct atgttctctt gctttgtca gcaagatagc cagatcaatg	2820
tcgatcgtgg ctggctcga gatacctgca agaatgcat tgcgctgcca ttctccaaat	2880

-continued

tgcagttcgc	gcttagctgg	ataacgccac	ggaatgatgt	cgtegtgcac	aacaatggtg	2940
acttctacag	cgcggagaat	ctcgtctct	ccaggggaag	ccgaagtttc	caaaaggctg	3000
ttgatcaaa	ctcgcgcgt	tgtttcatca	agccttacgg	tcaccgtaac	cagcaaatca	3060
atatcaactg	gtggcttcag	gocgccatcc	actgcccagc	cgtacaaatg	tacggccagc	3120
aacgtcgggt	cgagatggcg	ctcgtatgac	ccaactacct	ctgatagttg	agtcgatact	3180
tcggcgatca	ccgcttcctc	catactcttc	ctttttcaat	attattgaag	cattttatcag	3240
ggttattgtc	tcatgagcgg	atacatattt	gaatgtattt	agaaaaataa	acaaatagct	3300
agctcaactg	gtcgtacgc	tccgggcgtg	agactgcccg	gggcgctgcg	gacacataca	3360
aagttacca	cagattccgt	ggataagcag	gggactaaca	tgtgaggcaa	aacagcaggg	3420
ccgcgccggt	ggcgtttttc	cataggctcc	gccctcctgc	cagagttcac	ataaacagac	3480
gcttttcg	tgcactctgt	ggagccgtga	ggctcaacca	tgaatctgac	agtacggcgg	3540
aaacccgaca	ggacttaaac	atccccaccg	tttcggcggg	gtcgtccctc	cttgcgtctc	3600
cctgttcgga	ccctgcggtt	taccggatac	ctgttcgcgc	ttctccctt	acgggaagtg	3660
tggcgctttc	tcatagctca	cacactggta	tctcggtcgc	gtgtaggtcg	ttcgtccaa	3720
gctgggctgt	aagcaagaac	tccccgttca	gcccgactgc	tgcgccttat	ccggtaactg	3780
ttcaactgag	tccaaccogg	aaaagcacgg	taaaacgcca	ctggcagcag	ccattggtaa	3840
ctgggagttc	gcagaggatt	tgtttagcta	aacacgcggt	tgtctttgaa	gtgtgcgcca	3900
aagtccggct	acactggaag	gacagatttg	gttgcgtgtc	tctgcgaaag	ccagttacca	3960
ccggttaagca	gttccccaac	tgacttaacc	ttcgatcaaa	ccacctcccc	agggtggttt	4020
ttcgtttaca	gggcaaaaga	ttacgcgcag	aaaaaaagga	tctcaagaag	atcctttgat	4080
cttttctact	gaaccgctct	agatttccag	gcaatttacc	tcttcaaatg	tagcacctga	4140
agtcagcccc	atacgatata	agttgtaatt	ctcatgttag	tcatgccccg	cgcccaccgg	4200
aaggagctga	ctgggttgaa	ggctctcaag	ggcatcggtc	gagatccccg	tgccaatga	4260
gtgagctaac	ttacattaat	tgcgttgccg	tcaactgccg	ctttccagtc	gggaaacctg	4320
tcgtgccagc	tgcattaatg	aatcggccaa	cgcgcgggga	gaggcgggtt	gcgtattggg	4380
cgccaggggt	gtttttcttt	tcaccagtga	gacgggcaac	agctgattgc	ccttcaccgc	4440
ctggccctga	gagagttgca	gcaagcggtc	cacgctggtt	tgccccagca	ggcgaaaatc	4500
ctgtttgatg	gtggttaacg	gcgggatata	acatgagctg	tcttcgggat	cgtcgtatcc	4560
cactaccgag	atgtccgcac	caacgcgcag	cccggactcg	gtaatggcgc	gcattgcgcc	4620
cagcgcctac	tgatcgttgg	caaccagcat	cgcagtgga	acgatgccct	cattcagcat	4680
ttgcatgggt	tgttgaaaac	cggacatggc	actccagtcg	ccttccggtt	ccgctatcgg	4740
ctgaatttga	ttgcgagtga	gatatttatg	ccagccagcc	agacgcagac	gcgccgagac	4800
agaacttaat	gggcccgcta	acagcgcgat	ttgctgggtga	cccaatgcga	ccagatgctc	4860
cacgcccagt	cgcgtaccgt	cttcatggga	gaaaataata	ctgttgatgg	gtgtctggtc	4920
agagacatca	agaaataacg	ccggaacatt	agtgcaggca	gcttccacag	caatggcatc	4980
ctggtcatcc	agcggatagt	taatgatcag	cccactgacg	cgttgccgca	gaagattgtg	5040
caccgccgct	ttacaggctt	cgacgcgct	tcgttctacc	atcgacaacca	ccacgctggc	5100
accagttga	tcggcgcgag	atttaacgc	cgcgacaatt	tgcgacggcg	cgtgcagggc	5160
cagactggag	gtggcaacgc	caatcagcaa	cgaactgttg	cccgccagtt	gttgtgccac	5220

-continued

```

gcggttggga atgtaattca gctccgccat cgccgcttcc actttttccc gcgttttcgc 5280
agaaacgtgg ctggcctggg tcaccacgcg ggaaacggtc tgataagaga caccggcata 5340
ctctgcgaca tcgtataacg ttactggttt cacattcacc acctgaatt gactctcttc 5400
cgggcgctat catgccatac cgcgaaaggt tttgcgccat tcgatgggtg ccgggatctc 5460
gacgctctcc cttatgcgac tctcgcatta ggaaattaat acgactcact ata 5513

```

<210> SEQ ID NO 17

<211> LENGTH: 7118

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 17

```

gtcagcaagc tctggaatth cccgattctc tgatgggaga tccaaaaatt ctgcagctcc 60
ctcaatcacg atatcggctc tggatcgccc tgtagcttcc gacaactgct caattttttc 120
gagcatctct accgggcacg ggaatgaaat taacgggtgt ttagccatgt gttatacagt 180
gtttacaact tgactaacia atacctgcta gtgtatacat attgtattgc aatgtatacg 240
ctatthttcac tgctgtcttt aatggggatt atcgcaagca agtaaaaaag cctgaaaacc 300
ccaataggta agggattccg agcttactcg ataattatca cctttgagcg cccctaggag 360
gaggcgaaaa gctatgtctg acaaggggtt tgaccctga agtcggttgcg cgagcattaa 420
ggtctgcgga tagcccataa catacttttg ttgaaactgt gcgcttttat caaccctta 480
agggcttggg agcgttttat gcggccgcgg gggggggggg gaaagccacg ttgtgtctca 540
aaatctctga tgttacattg cacaagataa aaatatatca tcatgaacia taaaactgtc 600
tgcttacata aacagtaata caaggggtca tatgcacaat gaattgaaaa tcacggatat 660
gcaaacgctg gaaaccaaca ccaagacgac cgaagagtct attgacacca atagcctgaa 720
cctgccggac thtactaccg acagctacia ggatgcctat tctcgcatta acgccatcgt 780
tattgagggc gaacaggaag ctcatgacia ttacatctcc atcgcaacgc tgatcccga 840
tgagctggaa gagctgacga agctggcacg tatggagctg aaacacaaga aaggthttac 900
tgctgtcggg cgtaatctgg gtgtggacgc agacatggtt ttcgcgaaaa agthcttcag 960
caactgtcac ggcaatthcc aaatcgcgct ggaaaaaggt aacctgacca cctgcttgc 1020
gatccaagcg attctgatcg aagcatttgc gatttccgcy tacaatgtht acatccgtgt 1080
ggccgaccca thtgcaaaa agattaccga ggggtttgtc aaagacgagt atctgcatct 1140
gaactatggt caggagtggc tgaaaaagaa tctgtcccag tgtaaaagaag agctgatgga 1200
ggccaacaag gtcaatctgc cgtgattaa gaaaatgctg gacgaagtgg cagaagatgc 1260
gagcgtthtg gcgatggatc gtgaagagtt gatggaagag thcatgattg cgtaccagga 1320
tacctgttg gagattggcc tggataatcg cgaaattgcc cgtatggcga tggcggccat 1380
tgthtagtaa thtttctaath taactaataa aggaagtctg aatgthtggg ctgattggtc 1440
atagcaccag cthtgaggac gcaaaagcga aggcgagcct gctgggtthc gaccacatcg 1500
cggatggcga tctggatgtg tgggtgaccg caccgccgca actggttgaa aacgtggaag 1560
tcaaaagcgc gacgggtatc agcattgaag gttagctatat cgatagctgc thcgtgccg 1620
agatgctgag ccgcttaag accgcgcgct gtaaagthct gaatgcaatg gagctggcgc 1680
agaaaaaggg tatcaatath actgccttgg gtgctthtac thcattathc thtgagaact 1740

```

-continued

tcaacctggt	gcagcacaag	caaatccgta	ataccagcct	ggagtgggag	cgtttcacca	1800
cgggtaacac	gcacacggca	tgggtgattt	gtcgtcagct	ggagatcaac	gcaccgcgca	1860
ttggcatcga	cctgaaaact	gcaacggtcg	ctggtatcgg	cgcgaccggc	gatattggta	1920
gcgcggtgtg	tcgctggctg	gtcaataaga	ccggcattag	cgaactgctg	atggtcgctc	1980
gccaacaaca	gccactgacc	ctgctgcaaa	aagaactgga	cggtggcacc	atcaagagcc	2040
tggatgaagc	cctgccgcag	gcggatattg	tcgtgtgggt	tgettccgatg	cctaagacga	2100
tcgaaattga	gattgaaaac	ctgaaaaagc	cgtgcctgat	gatcgacggg	ggctaccoga	2160
agaatctgga	cgagaaatc	aaaggcaaaa	acattcacgt	gttgaagggt	ggatcgtcgc	2220
agtttttcaa	cgacattggc	tggaaatga	tggagtggc	ggagatgcaa	aaccgcgagc	2280
gtgagatggt	tgcgtgcttc	gccgaagcta	tgattctgga	gtttgagaaa	tgccatacca	2340
actttagctg	gggccgtaac	aatatcagct	tggagaagat	ggagttcac	ggtgctgcat	2400
ctctgaagca	cggtttcagc	gcgatcggtc	tggataaaca	gccgaaagtc	ttgaccgttt	2460
gaaattgaat	tcggttttcc	gtcctgtctt	gattttcaag	caaacaatgc	ctccgatttc	2520
taatcggagg	catttgtttt	tgtttattgc	aaaaacaaaa	aatattgtta	caaattttta	2580
caggctatta	agcctaccgt	cataaataat	ttgccattta	ctagttttta	attaaccaga	2640
accttgaccg	aacgcagcgg	tgtaacggc	gcagtggcgg	ttttcatggc	ttgttatgac	2700
tgtttttttg	gggtacagtc	tatgcctcgg	gcataccaagc	agcaagcgcg	ttacgccgtg	2760
ggtcgatggt	tgatgttatg	gagcagcaac	gatgttacgc	agcagggcag	tcgccctaaa	2820
acaaagttaa	acatcatgag	ggaagcgggt	atcgccgaag	tatcgactca	actatcagag	2880
gtagttggcg	tcatcgagcg	ccatctcgaa	ccgacgttgc	tggccgtaca	ttgtacggc	2940
tccgcagtg	atggcggcct	gaagccacac	agtgatattg	atttgcgggt	tacggtgacc	3000
gtaaggcttg	atgaaacaac	gcggcgagct	ttgatcaacg	accttttggg	aacttcggct	3060
tcccctggag	agagcgagat	tctccgcgct	gtagaagtca	ccattgttgt	gcacgcagac	3120
atcattccgt	ggcgttatcc	agctaagcgc	gaactgcaat	ttggagaatg	gcagcgcaat	3180
gacattcttg	caggatatct	cgagccagcc	acgatcgaca	ttgatctggc	tatcttgctg	3240
acaaaagcaa	gagaacatag	cgttgccctg	gtaggtccag	cgccggagga	actctttgat	3300
ccggttccctg	aacaggatct	atgtgagcgg	ctaaatgaaa	ccttaacgct	atggaactcg	3360
ccgcccgaact	gggctggcga	tgagcgaat	gtagtgccta	cgttgtccc	catttggtac	3420
agcgcagtaa	ccggcaaaat	cgcgccgaag	gatgtcgctg	ccgactgggc	aatggagcgc	3480
ctgccggccc	agtatcagcc	cgtcactact	gaagctagac	aggcttatct	tggacaagaa	3540
gaagatcgct	tggcctcgcg	cgcagatcag	ttggaagaat	ttgtccacta	cgtgaaaggc	3600
gagatcacca	aggtagtcgg	caaataatgt	ctaacaatc	gttcaagccc	acgccgcttc	3660
gcggcgcggc	ttaactcaag	cgttagatgc	actaagcaca	taattgctca	cagccaaact	3720
atcaggctcaa	gtctgctttt	attattttta	agcgtgcata	ataagcccta	cacaaattgg	3780
gagatatact	atgaggcgcg	ccacgagtgc	ggggaattt	cgggggcgat	cgcccctata	3840
tcgcaaaaag	gagttacccc	atcagagcta	tagtcgagaa	gaaaaccatc	attcaactcaa	3900
caaggctatg	tcagaagaga	aactagaccg	gatcgaagca	gccctagagc	aattggataa	3960
ggatgtgcaa	acgtcccaaa	cagagcttca	gcaatcccaa	aatggcagg	acaggacatg	4020
ggatgttggtg	aagtgggtag	gcggaatctc	agcgggccta	gcggtgagcg	cttccattgc	4080
cctgttcggg	ttggtcttta	gattttctgt	ttccctgcca	taaaagcaca	ttcttataag	4140

-continued

tcatacttgt	ttacatcaag	gaacaaaaac	ggcattgtgc	cttgcaaggc	acaatgtctt	4200
tctcttatgc	acagatgggg	actggaacc	acacgcacaa	ttcccttaa	aagcaaccgc	4260
aaaaataac	catcaaaata	aaactggaca	aattctcatg	tgggocggcc	aaaatgaagt	4320
gaagtcccta	tactttctag	agaataggaa	cttctatagt	gagtcgaata	agggcgacac	4380
aaaatttatt	ctaaatgcat	aataaatact	gataacatct	tatagtttgt	attatatttt	4440
gtattatcgt	tgacatgtat	aattttgata	tcaaaaactg	attttccctt	tattattttc	4500
gagatttatt	ttcttaattc	tctttaacaa	actagaaata	ttgtatatac	aaaaaatcat	4560
aaataataga	tgaatagttt	aattataggt	gttcatcaat	cgaaaaagca	acgtatctta	4620
tttaaagtgc	gttgcttttt	tctcatttat	aaggttaaat	aattctcata	tatcaagcaa	4680
agtgacaggc	gcccttaaat	attctgacaa	atgctctttc	cctaaactcc	ccccataaaa	4740
aaaccgcgcg	aagcgggttt	ttacgttatt	tgccgattaa	cgattactcg	ttatcagaac	4800
cgcccagggg	gcccagcctt	aagactggcc	gtcgttttac	aacacagaaa	gagttttag	4860
aaacgcaaaa	aggccatocg	tcaggggctt	tctgcttagt	ttgatgcctg	gcagttccct	4920
actctcgctt	tccgcttcc	cgctcaactga	ctcgctgcgc	tcggtcgttc	ggctgcggcg	4980
agcggtatca	gctcactcaa	aggcggtaat	acggttatcc	acagaatcag	gggataacgc	5040
aggaaagaac	atgtgagcaa	aaggccagca	aaaggccagg	aaccgtaaaa	aggccgcggt	5100
gctggcgctt	ttccataggc	tccgcccccc	tgacgagcat	cacaaaaatc	gacgctcaag	5160
tcagaggtgg	cgaaaccoga	caggactata	aagataccag	gcgtttcccc	ctggaagctc	5220
cctcgctgcg	tctcctgttc	cgaccctgcc	gcttaaccga	tacctgtccg	cctttctccc	5280
ttcgggaagc	gtggcgcttt	ctcatagctc	acgctgtagg	tatctcagtt	cggtgtaggt	5340
cgttcgctcc	aagctgggct	gtgtgcacga	acccccgctt	cagcccgacc	gctgcgcctt	5400
atccggtaac	tatcgtcttg	agtccaaccc	ggtaagacac	gacttatcgc	cactggcagc	5460
agccactggt	aacaggatta	gcagagcgag	gtatgtaggc	ggtgctacag	agttcttgaa	5520
gtggtgggct	aactacggct	acactagaag	aacagtattt	ggtatctgcg	ctctgctgaa	5580
gccagttaac	ttcggaaaaa	gagttggtag	ctcttgatcc	ggcaaacaaa	ccaccgctgg	5640
tagcggtggt	ttttttgttt	gcaagcagca	gattacgcgc	agaaaaaag	gatctcaaga	5700
agatcctttg	atcctttcta	cggggtctga	cgctcagtg	aacgacgcgc	gcgtaactca	5760
cgtaagggga	ttttggtcat	gagcttgcgc	cgctccgctc	agtcagcgta	atgctctgct	5820
tttaccatg	cttaatcagt	gaggcaccta	tctcagcgat	ctgtctattt	cgttcatcca	5880
tagttgcctg	actccccgct	gtgtagataa	ctacgatacg	ggagggctta	ccatctggcc	5940
ccagcgcctg	gatgataccg	cgagaaccac	gctcaccggc	tccggattta	tcagcaataa	6000
accagccagc	cggaaaggcc	gagcgcagaa	gtggtcctgc	aactttatcc	gcctccatcc	6060
agtctattaa	ttgttgccgg	gaagctagag	taagtagttc	gccagttaat	agtttgcgca	6120
acgttgttgc	catcgctaca	ggcatcgctg	tgctacgctc	gtcgtttggg	atggcttcat	6180
tcagctccgg	ttcccaacga	tcaaggcgag	ttacatgatc	ccccatggtg	tgcaaaaaag	6240
cggtagctc	cttcggctct	ccgatcgttg	tcagaagtaa	gttgcccgca	gtgttatcac	6300
tcatggttat	ggcagcaactg	cataattctc	ttactgtcat	gccatccgta	agatgctttt	6360
ctgtgactgg	tgagtactca	accaagtc	tctgagaata	gtgtatgcgg	cgaccgagtt	6420
gctcttgccc	ggcgtcaata	cgggataata	ccgcgccaca	tagcagaact	ttaaaagtgc	6480

-continued

tcacattg	aaaacgttct	tcggggcgaa	aactctcaag	gatcttaccg	ctgttgagat	6540
ccagttcgat	gtaaccact	cgtagacca	actgatcttc	agcatctttt	actttcacca	6600
gcgtttctgg	gtgagcaaaa	acaggaaggc	aaaatgccgc	aaaaaggga	ataaggcgca	6660
cacggaaatg	ttgaatactc	atattcttcc	ttttcaata	ttattgaagc	attatcagg	6720
gttattgtct	catgagcgga	tacatatttg	aatgtattta	gaaaaataa	caaatagggg	6780
tcagtgttac	aaccaattaa	ccaattctga	acattatcgc	gagccattt	atacctgaat	6840
atggctcata	acacccttg	ttgcctggc	ggcagtagcg	cggtggccc	acctgacccc	6900
atgccgaact	cagaagtga	acgcctagc	gccgatgga	gtgtggggac	tcccatgcg	6960
agagtaggga	actgccaggc	atcaataaaa	acgaaaggct	cagtcgaaag	actgggcctt	7020
tcgccgggc	taattatggg	gtgtgcacct	tattcgaact	tatagtgaag	ttcctattct	7080
ctagaaagta	taggaacttc	tgaagtgggg	cctgcagg			7118

<210> SEQ ID NO 18

<211> LENGTH: 13852

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 18

tgggagtcaa	taaaccgat	gtgcgttga	ttgccaacta	ccagccccc	ctgcaactca	60
gtgaatatct	ccaagagggtg	ggcgcgctg	ggcagatgg	cgaagcggca	caggccctgg	120
ttttggtgag	cgatcgctgg	ggcttgatc	gcgaagatca	acagcgttg	tcttttttc	180
agcaccaaag	tcaagacacc	tacaatcgcg	ccatggcact	tcagacgcag	ctgccctcc	240
agggtaatct	gcagcaactg	cggcaacact	ttcctgaagt	ggaattgacc	ctggcattac	300
tgcacaaaca	gggggccc	cgtggcaag	atcccttca	ctattgccgt	caacccttgg	360
cacaggtgcc	acccccacc	aaagaccctc	aagaacagtt	gatgcaaaag	ttcctctatc	420
accggggctg	ccgctggcag	ttctcctcc	aagccttgg	ttttgccact	gaggcaagg	480
gattccactg	tggccattgc	gatcgctgc	ggccgcgca	ccgctcccgc	aaaataccgt	540
aaattgccag	cgctgatca	ctggaatatt	gggtacactg	gcacatagaa	cggtcgcttt	600
accattggta	ggcaaaagt	tctcagcagt	cattctgttg	ccgcaaggta	ggggttgcag	660
gcattgggct	actacaagt	gaggaaatc	gcgaagcact	tcaagatgtg	ctttcagaac	720
acgcccttgt	tgtgcaagt	aatcagtttc	gcaaccaatt	aaacattatt	ttgaacaagc	780
cccccgccac	cgttgcccat	tattctgcc	tagcggattt	tctcaagtcg	cgcttgggac	840
agtttcatct	caatgatatt	gaccgcatta	aaataattgg	ccgcatacag	ggttcgccta	900
aaccggattg	ggaagaggtc	attgatctac	gtccccccaa	cccagcccta	gctgccctg	960
tgtatgcttc	ttctgccccg	tgggtgggtg	cgatcgctgc	tggctttgtc	agtttactgg	1020
tgatctttag	ctatcacctt	ggctcagtagc	agcaacagca	acggctgtag	ccgttgatcg	1080
aaggttcctt	tggcaaaaag	ggcgtcgtga	tgacggactt	taagtggcac	attgaggggtg	1140
gtacaggggtt	tattgtcggg	gttcttaaaa	actacagtaa	agggtatttt	cgcttagttc	1200
aggcggactt	tgaactcttt	gaccaaggcg	gtcagcaagt	tgggacagtg	gcggtacagg	1260
tttatggtct	tggccctgag	gaaacatggc	aattccgtga	actgatagcc	aatcatcagg	1320
cagtcgcagc	acggctggta	aaattacagt	cattcaatta	aggtttttct	aatgttttag	1380

-continued

ttccccagc	agggagcgac	accgcttgct	atggcacacc	ttaaagccct	gatctttgat	1440
gtcgatggca	ccttagcaga	tacggagcgg	gatggccatc	gtatgcctt	caacaaggcc	1500
tttgccgcg	ctggtctaga	ttgggaatgg	gacattcccc	tctatggtea	actcctggcg	1560
gtggctgggg	gcaaggagcg	gatccgggat	tacctgagt	gctttcgtec	cgattggcca	1620
cgccccaaa	atttggatgc	tctgattgcc	gatttacaca	aggccaagac	ccgctattat	1680
accgagctat	tggcggcagg	ggctattccc	ctgcggccgg	gggtgaaacg	gctcctcact	1740
gaagcccggg	aagcaggatt	acgtttggcg	atcgccacca	cgaccacccc	tgccaatgtc	1800
accgcactcc	ttgaaaatgc	cctcgctcct	gatggcgtea	gttggtttga	gataattgct	1860
gcccgggatg	tagttccagc	caagaaaccc	gcgcccgaca	tttacttcta	cacgcttgaa	1920
aagatgcgoc	tctcaccoca	agagtgcctt	gcctttgagg	attccgcaa	tgggattcag	1980
gcggccaactg	ccagtcacct	agcgaccatt	atcacgatta	ccgactacac	caaggatcat	2040
gattttcgtg	atgcagcgtc	ggtcctggat	tgcttagggg	aaccggacta	cccccttcag	2100
gttctgcgcg	gtgaggtggg	ttggacaacc	tatgtggatg	tccccctatt	gcgatcgctg	2160
caccagcagt	ggacaagcac	gttgagtcag	ggataatfff	ctggccgcag	cgttttacat	2220
tgaatatgac	ccccttagtc	tgaggatcaa	ggaacataat	gtacacgatt	gatttaattc	2280
tgcgatcatg	ccccatgccc	gtcagcattg	aacgcaagga	aagtgcagca	gcgatggcag	2340
tctatcagca	aattcagcag	gccatggcca	gtggtactcc	aactttcctc	gaactgacgt	2400
gcgatcgcca	agtgggcaag	aagttaacgg	tgctcacctc	agaaattgtc	gcccgtgcaaa	2460
tggcgataaa	ggatgcccc	tccagtaacta	tcagtctgtg	gggattcttt	gctcaattag	2520
tgagcaaac	cagcaactga	gggaaaatgc	ctcaataaag	ttgagttttt	cttggcaatg	2580
ctgattcttt	gccgttagga	tactaagcag	accgatccgt	aggggaaactg	gaagcaaatc	2640
ctccccgtct	gaaagtcagg	tatctctggt	gtgtcgtaat	agggttgtct	atggtgcagc	2700
gtttcctgcc	ggttctgatt	ttgttgggg	gtagttttgg	tcttgcgacc	cctgcccctg	2760
tgcgatgcca	agccaatcag	ggctttacgt	ttacttgggg	tgaggggccc	agtggccgac	2820
agcagttgca	ataccactta	gataacggca	cccccggttt	tatggcgcat	cgctattggc	2880
tgccgctggg	tcagcagaaa	gtggccatca	atcgcaataa	cattacctat	cccgaactact	2940
acaacgggat	tattgatccc	aaaggcattg	aggtgcgcat	cggtggcgat	cgccggcaatc	3000
gcttcttcca	atctcgcgt	gaccccgcca	ccaaaattca	attggcgga	gtctccgctg	3060
atcgcgataa	ccgcgtgatt	gatattgtgc	cggtgaggt	gattcccgc	ggaacaccgg	3120
tgcaagttat	tctcaataat	gtgcgcaacc	ctaacaatgg	cgccatgtac	tatttcaatg	3180
cccgcattgg	ctcccctgga	gatattcccc	tcatgcgcta	cgttggcacc	tgattctca	3240
gcattgcaaa	taactaaaac	ccgtcaaaact	cgagcattgg	tgagcgggtt	agccatttct	3300
aactattgcg	ggcgatcgc	cctagactag	ttttttgtct	attattgccc	gttcaactct	3360
tacaccagat	gccagattcc	gttaggtctt	cattcccctc	catttctcct	ctgctcacgc	3420
ctctgatgta	ccgcctcgtg	ggggaacttg	tcctgcggcg	ctatcttctg	acccttgagg	3480
tgcaagggca	ggagcgggtg	ccccaaaggg	gtccagtgat	cttggcccc	accacccgtt	3540
cccgtcggga	tgcctgatt	attcccctatg	tcactgggcg	gcgggtgagt	gggcgcgacc	3600
tctactacat	ggtgtcccac	gatgagatgt	tgggactaca	gggtgggtg	attgctcagt	3660
gtggcggttt	tcccgtcaat	acccaagcgc	cttcgggtgag	tgcttgcgt	acgggtgtgg	3720
aactgctcgg	gcaggggcaa	gccttggtgg	tgttccctga	ggggaatc	tttcgcatc	3780

-continued

gccagattca	tcccccaag	ccggggttgg	ctcgettagc	ccttcaggcg	gcccagcgc	3840
gtgaacaagc	aatccagatt	ctgccaattt	tactcgatta	tgcccagccc	taccacagc	3900
ggggaagtgc	ggccaaggta	atcattgggg	ctcccttgag	taccgacaat	tacgatgcca	3960
gccggccaaa	aagtgctgcc	caacaactga	ccagtgatct	ctttagaaga	cttcagcagc	4020
tccaaggggg	gcatcgccc	ctgtgttttg	cttagacctc	aaacttccat	ccccgcggcc	4080
gctcttgata	acccaagagg	gcatttttta	ggcgcgcctc	gagtaacacc	gtgctgtttg	4140
actattttac	ctctggcggg	gataatggtt	gcaggatcct	tttgctggag	gaagaattca	4200
tgacaacggc	taccgctaca	cctgttttgg	actaccatag	cgatcgctac	aaggatgcct	4260
acagccgc	taacgccatt	gtcattgaag	gtgaacagga	agctcacgat	aactatatcg	4320
atthagccaa	gctgctgcca	caacaccaag	aggaactcac	ccgccttgcc	aagatggaag	4380
ctcgccacaa	aaaggggttt	gagcctgtg	gtcgcaacct	gagcgtaacg	ccagatatgg	4440
aatttgccaa	agccttcttt	gaaaaactgc	gcgctaactt	tcagagggct	ctggcggagg	4500
gaaaaactgc	gacttgtcct	ctgattcaag	ctttgatcat	cgaatccttt	gcatcgccgg	4560
cctacaacat	ctacatccca	atggcggatc	ctttcgccc	taaaattact	gagagtgttg	4620
ttaaggacga	atacagccac	ctcaactttg	gcgaaactcg	gctcaaggaa	cactttgaaa	4680
gcgtcaaagg	agagctcgaa	gaagccaatc	gcgccaat	acccttggtc	tggaaaatgc	4740
tcaaccaagt	ggaagcagat	gccaaagtgc	tcggcatgga	aaaagatgcc	cttgtggaag	4800
acttcatgat	tcagtacagt	ggtgccttag	aaaatcgg	ctttaccacc	cgcgaaatta	4860
tgaagatgtc	agtttatggc	ctcactgggg	cataatggtg	gcttaacgta	tcgttacatt	4920
tcagtcacca	cacgcttgta	tgtttggatt	aattggtcat	ctgacgagtc	tggagcacgc	4980
ccaagccggt	gcccatcagt	tgggttacc	cgaatatgcc	gatcaaggct	tggaattttg	5040
gtgtatggca	ccgccgcaga	tcgtcgatga	gattacggtg	acgagcgtaa	cgggcaaac	5100
tatctatggc	aaatcgttg	agtcctgctt	tttaccagag	atgctggcca	accagcgggt	5160
gaaggcagcg	actcgcaag	ttattaacgc	catggcccat	gccccaaagc	acaacattga	5220
cattacggcc	ttggggggct	tctcctcgat	catctttgag	aactttgatc	tggagaaaat	5280
gtcccacatt	cgcaacattg	aactggactt	tcgccgcttt	acaacgggga	ataccatac	5340
egcctatc	atctgccaac	aaattgagca	ggcggcgc	caagtgggga	ttgatttgcg	5400
gcaggcaacc	gtggctgttt	gtggggctac	gggggatatt	ggtagtgccg	tctgccgttg	5460
gttgaatacc	tgtttagatg	tgcaagatct	cttactcgta	gcacggaatc	gcatcgccct	5520
gctggagcta	caggcggaat	tgggacgggg	gaaaatcctc	gacttgatgg	aggcgcctgcc	5580
ccttgccgat	attgtggttt	gggtggccag	tatgcccag	ggagttgagc	tgagcattga	5640
gcagttaaaa	cgcccctccc	tgatgattga	tgggtggtat	ccccaaaata	tggccaccaa	5700
aattcagcac	ccccagattc	atgttctcaa	tgggtgcat	gtcagcatg	ccctcgacat	5760
tgactgga	attatgga	ttgtgaatat	ggatgtgcc	tcgcggcaga	tgtttgctg	5820
ttttgcagag	gctatgcttt	tagagttega	ggctggcac	accaatttct	cttggggacg	5880
caatcaaatc	actgtgga	agatgcagca	aattggtgag	gtctcccgta	aacatggatt	5940
tcagccacta	ctgttgaatc	ctcagtaagc	ggccgcaaaa	aaaacgggcc	ggcgtattat	6000
cgccggcccg	agtaacaccg	tcgctgttga	ctattttacc	tctggcggtg	ataatggttg	6060
caggatcctt	ttgctggagg	aaaacatata	gaaaggacca	ataataatga	ctagagaaga	6120

-continued

aagaatgaag attgttcatg aaattaagga acgaatattg gataaatatg gggatgatgt	6180
taaggcaatt ggtgtttatg gctctcttgg tcgctcagact gatgggccct attcggatat	6240
tgagatgatg tgtgttctgt caacagaggg agtagagttc agctatgaat ggacaaccgg	6300
tgagtggaag gcggaagtga atttttatag cgaagagatt ctactagatt atgcatctcg	6360
ggtggaaccg gattggccgc ttacacatgg tcgatttttc tctattttgc cgatttatga	6420
tccaggtgga tactttgaga aagtgtacca aactgctaaa tcggtagaag cccaaaagt	6480
ccacgatgcg atctgtgccc ttatcgtaga agagctgttt gaatatgcag gcaaatggcg	6540
taatattcgt gtgcaaggac cgacaacatt tctaccatcc ttgactgtac aggtggcaat	6600
ggcaggtgcc atgttgattg gtctgcatca tcgcatctgt tatacgacga gcgcttcggt	6660
cttaactgaa gcagttaagc aaccagatct tcctccaggt tatgtccaac tgtgccagct	6720
cgtaatgtct ggtcaacttt ccgaccctga gaaacttctg gaatcgctag agaatttctg	6780
gaatggggtt caggagtggtg cggaaacgaca cggatatata gtggatgtgt caaaacgcat	6840
accattttga tgtctaacc ccttccttgc ccacagcttc gtcgatggcg cgaaatttgc	6900
ggtaaatata atgaccctct tgataacca agagggcatt ttttaggcgc gccctaagcg	6960
tccgtaggca caattaaggc ttcaaattgt tggcgaagct gctcagtcac ttccttgacg	7020
gcttgccgtg ccccttggcg atcgcgccgg tacagagggc aatagctctc taaattgaga	7080
gggtcgccga cactgaggcg cacctgccgc aaaccacca aacgattgag attcagcctt	7140
tttccctcta gccaatcaaa tgtgcgccag agaatcagcg cgacatctgc aaagcgatga	7200
atcgtgaatt tctcacggat atagctacc gtaattgagg taaatcgctc cgcaagacgc	7260
atatgacgca atcgcacatt ggcttctctg gccaaccaat cggctaggca gcgctctacg	7320
gccgaaagt gtgccaaatc actgcgaaac atccgttccc aagcagcctg ttcfaatcgt	7380
cggcagcgac tcacaaaatc ggcaactggc ttcagaccaa agtaggactc tgccaccaca	7440
agggcgctgt tgaggaggcg ctgaattcgc gctgccaat tagcattggc agagtcaaag	7500
gggggcagtt cgggaaaatc ttgaccatag gaggtggcat aaaaagcctc caggcgatcc	7560
aagaggtgga tcgctaaatt cagcagggcg cggtagaggt cgtctggctg ggtactgtga	7620
gaatctgtag ggcacccaag gcggttctcc agttgtgcca tcagccttgc catgcctcc	7680
caagagggt gactgaggct gtactgaatg ccaatgggaa gaatgaccac ggggagcgat	7740
cgccccgcct tggctaaatc ttctagacac caaaatccca gttgggccac ccccgctcc	7800
aaagtgcgga ccagttcgtt gtgctcattc gttgctccct ccggcgctgc cgctagggga	7860
aatcgtctc cgagaagtag ctcccgctc gagcgcaggg cttggctatc gagcttaccg	7920
cgcatgatgg aaatcccccc caaccgtgaa aagagccaac caatctgcgc cctgcccag	7980
aggggaatcc cgcgatcgta gagaaaatag ccatttgcg gcggacgcaa gggaatgccc	8040
agccgccgtg ctgtttgcgg cagtaaatgc cacatcaat agcccatcac caacggatca	8100
tccgtacagg gatggcgaaa ggcaatgagg agccggacct gtccctgctg aaactgctgg	8160
taataacggg caagggtctc cacattcacc ccttcaacct gctgtagccc aagaccatag	8220
cgaatgtaga ggggcaggag tcttgctact gtccaccaga cgggtagct aaaccgctgg	8280
gggagaaaat gcaacggcgg ttgggcagtt gtcactacac tggacattag gcaagctcct	8340
cagggcaatg gctaaactga ggcagtggcc aactccgcaa ttaactgctc taacatcgtt	8400
tgatcggccc aatagacagc attacaaaac tgacaggtgg cttctgcctt tgcctctgtg	8460
gctaggat at ctcttaattc tgctccctc aggagcttga gtgccgctaa catccgttca	8520

-continued

tgggaacagc	cacagtggaa	gcgaccatt	tgcggttggg	gcaagatttg	taaatccata	8580
tcccctaaga	gttcctgaaa	gatatctggc	agtgtccgcc	ctgcctgtag	cagtgggtga	8640
aagcccttaa	gattggccac	ccgttgttca	agggtcgcga	tcagggttcc	atcattggcc	8700
gctttgggta	gcacctgtaa	catcaacca	ccggcggcag	tcaccccgga	ctcttcgaca	8760
aaaaacacca	acatcagggc	ggaggggggt	tgctctgagg	tggcagagga	gtaggtgatg	8820
tcttctgcaa	tttcgcccga	gactagctcc	accgtgctgg	aataggggta	gcccagtagcca	8880
agatcgtgga	tgacgtagag	atatccctga	tggcccacgc	ctgccccac	atcgagtttg	8940
cccttggeat	tggggggcag	ttcaacctg	gggtaactgca	catagccgcg	aactgtgcca	9000
tgcgcaccag	catcggcaaa	aatggttcct	aggggaccgt	tgccctgaat	gcgcacattc	9060
accctgtcct	ggggctgttt	gaaactggag	gcaaggatta	agcctgcggc	catggttcgt	9120
cccaaggcgc	ctgtggccac	gtaggacagt	tggtagcgtt	tgcgggcttc	atcagtgagt	9180
tgagtggtaa	tcacacctac	ggcccggatg	ccttcggcag	cggcagttgc	tcgcaacaga	9240
aaatcggcca	tgttcaacct	acgaaatgtt	ttgttacatt	tagtgtgaca	tactcccacc	9300
gctgaccagg	gcacaatggg	gcaaaaaacc	atcaatcctg	cctttggtga	ccgatccagt	9360
acagccagcc	agggtctaag	actgggaaga	cccctagcac	tggggctaga	aaattggcga	9420
tgataggcaa	gcaatagtea	ttcagcgtcc	agtcattccg	cctatggcca	tgcccctcac	9480
tgtcttgctt	gccacaactg	ttttgacaga	agcgactcaa	ttgccccagg	gcggcttgat	9540
tacggagatt	ccgacgctgg	cgatcgccca	ccgtttggcc	cagcagttgc	gcccgccattg	9600
gccccagag	acccccttaa	cgctgattga	tgcgcaatac	cagagtatcc	ccctgaccct	9660
tggggaattg	gcccagctca	ccgatgcca	ctgtccttta	cagctctatg	tgccgcccc	9720
cttgccagag	gccttgacgc	aatttcaacg	cctgatggat	gtggttcgag	agctgcgcca	9780
tccggagcgt	ggctgtcctt	gggatttgca	gcaaaoccca	accagtctca	ttccctatgt	9840
ccttgaggaa	gcctatgaag	tggtacatgc	cctgcaggag	ggagatgcgg	gggogatcgc	9900
cgaagaattg	ggagacctgt	tgcttcaagt	tgttctccag	agccaacttg	cccaagaagc	9960
cggccaattt	acccttgctc	aagtcattca	aaggattacc	gataaactca	tccgcccga	10020
tccccacgtc	tttgggtaag	tggcactcac	caactgctca	gaggtgcgcg	accaatggga	10080
gcaaatcaaa	gcgctgaaa	aaggcaccga	actccccctg	agtcaaacgc	tgcaacgtta	10140
cgcaacgacc	ctcccacccc	tgatggccgg	catgaaaatt	ggtgagcgcg	ccagtcgcgc	10200
tggcctcgat	tggccgacga	ttagtgggtc	atgggagaaa	ttttacgagg	aactggcgga	10260
gtttcaggag	gccccttctgc	aagggaatgc	tgagcaacag	gcagcggaat	taggagacct	10320
gctcttcagt	gtgattaacc	ttgcccctg	gtgccaactg	gatcctgtta	atgccctgca	10380
acaaacctac	caacgcttta	ttcaacgctt	ggcctgtatt	gagcagctca	tcgatcgcgc	10440
ccttgagacg	tacaccctag	aagaactaga	agcccctctg	caacaggcca	aagtacagtt	10500
agccaccgac	agcagggcaa	cccctatgga	gactgaggaa	gaggcctagt	ccgctgcggc	10560
ccttgccacc	ttcagttcat	cgagattcca	cagggggccc	cccagcgcgc	tgggcttggc	10620
gccaatgaca	tgattgcgaa	aagctgtaag	ggagagggga	ttcacgaggt	aaataaagg	10680
gagatattcc	tgagctagtc	gttgggcttc	cgcataaatt	tgctgcccgc	gttccagatt	10740
gagctcctgg	gcaccttggg	catacaggtc	actgatgcgc	tgctcccagt	cagcgcgcgc	10800
tcgacccgta	atgggtgggt	gattcgggtg	cggttgctga	ttgaatgtat	gcaaaaggcc	10860

-continued

atccacacgc cagatattgg caccgctatt gggttcattg cccccccag taaagccgag 10920
gatatgggct tcccactcta ggggaattgga gagacgatcc acgagggtag caaaggccaa 10980
aaattgcaga tccacctgca tgcgatcgc ccttaggtcc tgetgaactt ggcgcgggcc 11040
ggccaaaatg aagtgaagt cctatacttt cttagagaata ggaacttcta tagtgagtcg 11100
aataagggcg acacaaaatt tattctaaat gcataataaa tactgataac atcttatagt 11160
ttgtattata ttttgatta tctgtgacat gtataatfff gatataaaa actgattttc 11220
cctttattat tttcgagatt tttttctta attctcttta acaaactaga aatattgtat 11280
atacaaaaaa tcataaataa tagatgaata gtttaattat aggtgttcat caatcgaana 11340
agcaacgat cttatttaaa gtgcgttgc tttttctcat ttataagggt aaataattct 11400
catatatcaa gcaaagtgc aggcgcctt aaatattctg acaaatgctc tttccctaaa 11460
ctcccccat aaaaaaacc gccgaagcgg gtttttactg tatttgccga ttaacgatta 11520
ctcgttatca gaaccgccca gggggcccca gcttaagact ggccgtcgtt ttacaacaca 11580
gaaagagtt gtagaaacgc aaaaaggcca tccgtcaggg gccttctgct tagtttgatg 11640
cctggcagtt ccctactctc gccttcogct tectcctca ctgactcgtc gcgctcggtc 11700
gttcggctgc ggcgagcgg atcagctcac tcaaaggcgg taatacgggt atccacagaa 11760
tcaggggata acgcaggaana gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt 11820
aaaaaggcgg cgttgctggc gtttttccat aggtccgcc cccctgacga gcatcacaana 11880
aatcagcgt caagtcaag gtggcgaaac ccgacaggac tataaagata ccaggcgtt 11940
ccccctgga gctccctcgt gcgctctct gttccgacct tgccgcttac cggatactg 12000
tccgccttcc tccctcggg aagcgtggcg ctttctcata gctcacgctg taggtatctc 12060
agttcgggtg aggtcgttcg ctccaagctg ggcgtgtgac acgaaacccc cgttcagccc 12120
gaccgctcgg ccttatccgg taactatcgt cttgagttca acccggtaa acacgactta 12180
tcgccactgg cagcagccac tggtaaacag attagcagag cgagggtatg aggcgggtgct 12240
acagagttct tgaagtggg ggctaactac ggctacacta gaagaacagt atttggatc 12300
tgcgctctgc tgaagccagt tacctcggg aaaaagagtg gtagctctg atccggcaana 12360
caaaccaccg ctggtagcgg tggtttttt gtttcaagc agcagattac gcgcagaaaa 12420
aaaggatctc aagaagatcc tttgatctt tctacgggg ctgacgctca gtggaacgac 12480
gcgcgcgtaa ctcacgttaa gggattttg tcatgagct gcgcgctccc gtcaagtcag 12540
cgtaatgctc tgcctttacc aatgcttaat cagttaggca cctatctcag cgatctgct 12600
atctcgttca tccatagttg cctgactccc cgtcgtgtag ataactacga tacgggaggg 12660
cttaccatct gggcccagcg ctgcgatgat acccgagana ccacgctcac cggctccgga 12720
tttatcagca ataaaccagc cagccggaag ggcgagcgc agaagtggtc ctgcaacttt 12780
atccgcctcc atccagttc ttaattgtg ccgggaagct agagtaagta gttcggcag 12840
taatagttg cgcaacgtt tggcatcgc tacaggcctc gtggtgtcac gctcgtcgt 12900
tggtatggct tcattcagct ccggttccca acgatcaagg cgagttacat gatccccat 12960
gttgtgcaana aaagcggtta gctcctcgg tctccgactc gttgtcagaa gtaagttggc 13020
cgcagtgta tcaactcag ttatggcagc actgcataat tctcttactg tcatgccatc 13080
cgtaagatgc tttctgtga ctgggtgagta ctcaaccaag tcattctgag aatagttgat 13140
gcggcgaccg agttgctctt gcccgcgctc aatacgggat aataccgccc cacatagcag 13200
aactttaana gtgctcatca ttgaaaacg ttctcgggg cgaaaactct caaggatctt 13260

-continued

```

accgctgttg agatccagtt cgatgtaacc cactcgtgca cccaactgat cttcagcacc 13320
ttttactttc accagcgttt ctgggtgagc aaaaacagga aggcaaaatg cgcacaaaaa 13380
gggaataaagg gcgacacgga aatggtgaat actcatattc ttcctttttc aatattattg 13440
aagcatttat cagggttatt gtctcatgag cggatacata tttgaatgta tttagaaaaa 13500
taaacaaata ggggtcagtg ttacaaccaa ttaaccaatt ctgaacatta tcgagagccc 13560
attatacct gaatatggct cataacaccc cttgtttgcc tggcggcagt agcgcggtgg 13620
tcccacctga ccccatgcc aactcagaag tgaaacgcc tagcggccgat ggtagtgtgg 13680
ggactcccca tgcgagagta gggaaactgcc aggcacaaa taaaacgaaa ggctcagtcg 13740
aaagactggg cctttcggcc gggctaatta tggggtgtcg cccttattcg actctatagt 13800
gaagtcceta ttctctagaa agtataggaa cttctgaagt ggggcctgca gg          13852

```

```

<210> SEQ ID NO 19
<211> LENGTH: 92
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        oligonucleotide

```

```

<400> SEQUENCE: 19

```

```

gcgggcgcgc gagtaacacc gtgcgtgttg actattttac ctctggcggg gataatgggt      60
gcaggatcct tttgctggag gaaaaccata tg                                     92

```

```

<210> SEQ ID NO 20
<211> LENGTH: 605
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        polynucleotide

```

```

<400> SEQUENCE: 20

```

```

gcgggcgcgtt cgttataaaa taaacttaac aaatctatac ccacctgtag agaagagtcc      60
ctgaatatca aaatggtggg ataaaaagct caaaaaggaa agtaggctgt ggttccctag      120
gcaacagtct tccctacccc actggaact aaaaaaacga gaaaagttcg caccgaacat      180
caattgcata attttagccc taaaacataa gctgaacgaa actggttgtc ttccttccc      240
aatccaggac aatctgagaa tcccctgcaa cattacttaa caaaaaagca ggaataaaat      300
taacaagatg taacagacat aagtcccac accgttgat aaagttaact gtgggattgc      360
aaaagcattc aagcctagc gctgagctgt ttgagcatcc cgggtggccc tgcgctgcc      420
tccgtgtttc tccctggatt tatttaggta atatctctca taaatccccg ggtagttaac      480
gaaagttaat ggagatcagt aacaataact ctagggtcat tactttggac tccctcagtt      540
tatccggggg aattgtgttt aagaaaatcc caactcataa agtcaagtag gagattaatc      600
atattg                                           605

```

```

<210> SEQ ID NO 21
<211> LENGTH: 1510
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        polynucleotide

```

```

<400> SEQUENCE: 21

```

-continued

gcgcccgga aggcgaagcg gcatgcattt acgttgacac catcgaatgg tgcaaacct	60
ttcgcggtat ggcgatgatag cgcccgaag agagtcaatt cagggtgggtg aatatgaaac	120
cagtaacggtt atacgatgtc gcagagtatg ccggtgtctc ttatcagacc gtttcccgcg	180
tggatgaacca ggccagccac gtttctgcca aaacgcggga aaaagtggaa gcggcgatgg	240
cggagctgaa ttacattccc aaccgcgtgg cacaacaact ggccggcaaa cagtcgttgc	300
tgattggcgt tgccacctcc agtctggccc tgcaecgccc gtcgcaaatt gtcgcgcgca	360
ttaaatctcg cgccgatcaa ctgggtgcca gcgtgggtgt gtcgatggta gaacgaagcg	420
gcgtcgaagc ctgtaaagcg gcggtgcaca atcttctcgc gcaacgcgtc agtgggtga	480
tcattaacta tccgctggat gaccaggatg ccattgctgt ggaagctgcc tgcactaatg	540
ttccggcgtt atttcttgat gtctctgacc agacacccat caacagtatt atttctccc	600
atgaagacgg tacgcgactg ggcgtggagc atctggctgc attgggtcac cagcaaatcg	660
cgctgttagc gggccatta agttctgtct cggcgcgtct gcgtctggct ggctggcata	720
aatatctcac tcgcaatcaa attcagccga tagcggaaagc ggaagcgac tggagtcca	780
tgtccggttt tcaacaaacc atgcaaatgc tgaatgaggc catcgttccc actgcgatgc	840
tggttgcaaa cgatcagatg gcgctggcg caatgcgcgc cattaccgag tccgggctgc	900
gcgttggtgc ggatatctcg gtagtgggat acgacgatac cgaagacagc tcatgttata	960
tcccgcctgc aaccaccatc aaacaggatt ttcgcctgct gggcacaacc agcgtggacc	1020
gcttgctgca actctctcag ggcagggcgg tgaagggcaa tcagctgttg cccgtctcac	1080
tggatgaaaag aaaaaccacc ctggcgccca atacgcaaac cgctctccc cgcgcttg	1140
ccgattcatt aatgcagctg gcacgacagg tttcccgact ggaagcggg cagtgagcgc	1200
aacgcaatta atgtgagtta gcgcgaattg atctggtttg acagcttacc atcgagctcg	1260
actgcacggt gcaccaatgc ttctggcgtc aggcagccat cggaaagctgt ggtatggctg	1320
tgcaggtcgt aatcactgc ataattcgtg tcgctcaagg cgcactcccg ttctggataa	1380
tgttttttgc gccgacatca taacggttct ggcaaatatt ctgaaatgag ctgttgacaa	1440
ttaatcatcc ggctcgata atgtgtggaa ttgtgagcgg ataacaattt cacacaggaa	1500
acagcatatg	1510

<210> SEQ ID NO 22

<211> LENGTH: 78

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 22

gcgccgctg ttgacaatta atcatcgga tagtatatcg gcatagtata atacgacaag	60
---	----

gtgaggaact aacatatg	78
---------------------	----

<210> SEQ ID NO 23

<211> LENGTH: 5907

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 23

-continued

acaactcggc	ttccgagcct	ggctccacca	tggttatatc	tggagtaacc	agaatttcga	60
caacttcgac	gactatctcg	gtgcttttac	ctccaaccaa	cgcaaaaaca	ttaagcgcga	120
acgcaaagcc	gttgacaaag	caggtttata	cctcaagatg	atgaccgggg	acgaaattcc	180
cgcccattac	ttcccactca	tttatcgttt	ctatagcagc	acctgcgaca	aatttttttg	240
ggggagtaaa	tatctccgga	aacccttttt	tgaaacccta	gaatctacct	atcgccatcg	300
cgttgtctg	gccgcccgtt	acacgccaga	agatgacaaa	catcccgtcg	gtttatcttt	360
ttgtatccgt	aaagatgatt	atctttatgg	tcgttattgg	ggggcctttg	atgaatatga	420
ctgtctccat	tttgaagcct	gctattacaa	accgatccaa	tgggcaatcg	agcagggaa	480
tacgatgtac	gatecggggc	ctggcggaaa	acataagcga	cgacgtgggt	tcccggcaac	540
cccaaactat	agcctccacc	gtttttatca	accccgcag	ggccaagttt	tagacgctta	600
tattgatgaa	attaatgcc	tggagcaaca	ggaaattgaa	gcatcaatg	cggatattcc	660
ctttaaacgg	caggaagttc	aattgaaaat	ttcctagcct	cactagccaa	aagcgcgatc	720
gcccaccgac	catcctccct	tgggggagat	gcggccgctt	gtagcaattg	ctactaaaa	780
ctgcgatcgc	tgctgaaatg	agctggaatt	ttgtccctct	cagctcaaaa	agtatcaatg	840
attactta	gtttgtctg	cgcaaacctc	ttgcagaaca	tgcattgatt	acaaaagtt	900
gtagtttctg	ttaccaatg	cgaatcgaga	actgccta	ctgcccagta	tgcgactctt	960
tagcaggagg	aaaaccatat	gagatctgta	gtaggatccc	tcgagagtga	gagccggcga	1020
gctcatagta	tgtacatgat	gactgtaccc	atggttgaat	tcggttttcc	gtcctgtctt	1080
gattttcaag	caaacaatgc	ctccgatttc	taatcggagg	catttgtttt	tgtttattgc	1140
aaaaacaaaa	aatattgtta	caaattttta	caggctatta	agcctaccgt	cataaataat	1200
ttgccattta	ctagttttta	attaaccaga	accttgaccg	aacgcagcgg	tggtaacggc	1260
gcagtggcgg	ttttcatggc	ttgttatgac	tgtttttttg	gggtacagtc	tatgcctcgg	1320
gcatccaagc	agcaagcgcg	ttacgcctg	ggtcagatgt	tgatgttatg	gagcagcaac	1380
gatgttacgc	agcagggcag	tcgccctaaa	acaaagttaa	acatcatgag	ggaagcggtg	1440
atcgcgaag	tatcgactca	actatcagag	gtagttggcg	tcacgagcgc	ccatctcgaa	1500
ccgacgttgc	tggccgtaca	ttgtacggc	tccgcagtg	atggcggcct	gaagccacac	1560
agtgatattg	atttgcctgt	tacggtgacc	gtaaggcttg	atgaaacaac	gcggcgagct	1620
ttgatcaacg	accttttgg	aacttcggct	tcccctggag	agagcgagat	tctccgcgct	1680
gtagaagtca	ccattgttgt	gcacgacgac	atcattccgt	ggcgttatcc	agctaagcgc	1740
gaactgcaat	ttggagaatg	gcagcgcaat	gacattcttg	caggtatctt	cgagccagcc	1800
acgatcgaca	ttgatctggc	tatcttgctg	acaaaagcaa	gagaacatag	cgttgccttg	1860
gtaggtccag	cggcggagga	actccttgat	ccggttcctg	aacaggatct	atttgaggcg	1920
ctaaatgaaa	ccttaacgct	atggaactcg	ccgcccgact	gggctggcga	tgagcgaaat	1980
gtagtctta	cgttgtccc	catttggtag	agcgcagtaa	ccggcaaaat	cgcgccgaag	2040
gatgtcgtg	ccgactgggc	aatggagcgc	ctgccggccc	agtatcagcc	cgtcatactt	2100
gaagctagac	aggcttatct	tggacaagaa	gaagatcgtc	tggcctcgcg	cgagatcag	2160
ttggaagaat	ttgtccacta	cgtgaaaggc	gagatcacca	aggtagtcgg	caaaatattg	2220
ctaaacaattc	gttcaagcgc	acgccccttc	gcggcggcgc	ttaactcaag	cgttagatgc	2280
actaagcaca	taattgctca	cagccaaact	atcaggtcaa	gtctgctttt	attattttta	2340
agcgtgcata	ataagcccta	cacaaattgg	gagatata	atgaggcgcg	cctgatcagt	2400

-continued

tggtgctgca	ttagctaaga	aggtcaggag	atattattcg	acatctagct	gacggccatt	2460
gcatcataa	acgaggatat	cccactggcc	atthtcagcg	gcttcaaagg	caattttaga	2520
cccatcagca	ctaattgggtg	gattacgcac	ttcttggttt	aagttatcgg	ttaaattccg	2580
cttttgttca	aactcgcgat	catagagata	aatatcagat	tcgcccgcac	gattgaccgc	2640
aaagacaatg	tagcgaccaa	cttcagaaac	ggcaggatgg	gaggcaattt	catttagggt	2700
attgaggccc	ggtaacagaa	tcgtttgcct	ggtgctggta	tcaaatagat	agatatacctg	2760
ggaaccattg	cggtctgagg	caaaaacgag	gtagggttcg	gcatcgccg	ggtcaaatc	2820
gagggcccga	ctatttaaac	tcgcccacc	gggatcaacg	ggaaaatga	caatgcgcgg	2880
ataaccaacg	cagctctgga	gcagcaaac	gaggctaccg	aggaaaaaac	tcgctagaaa	2940
agaaacatag	cgcataggtc	aaagggaat	caaaggcggc	gcatcgcca	atthttctat	3000
aatattgtcc	taacagcaca	ctaaaacaga	gccatgctag	caaaaattg	gagtgcacc	3060
attgtcgggg	tcgatgcct	cagggtcggg	gtggaagtgg	atthttcggg	cggcttaccg	3120
aaaatgatgg	tggtcggact	gcggccggcc	aaaatgaagt	gaagttccta	tactthctag	3180
agaataggaa	cttctatagt	gagtcgaata	agggcgacac	aaaatttatt	ctaaatgcat	3240
aataaact	gataacatct	tatagtttgt	attatatttt	gtattatcgt	tgacatgtat	3300
aattttgata	tcaaaaactg	atthttccct	tattattttc	gagatttatt	ttcttaattc	3360
tctttaacaa	actagaaata	ttgtatatac	aaaaaatcat	aaataataga	tgaatagttt	3420
aattataggt	gttcatcaat	cgaaaaagca	acgtatctta	tttaagtgc	gttgctthtt	3480
tctcatttat	aaggttaaat	aattctcata	tatcaagcaa	agtgacagcg	gcccttaaat	3540
attctgacaa	atgctctttc	cctaaactcc	ccccataaaa	aaaccgcgcg	aagcgggttt	3600
ttacgttatt	tcggtgattaa	cgattactcg	ttatcagaac	cgcccagggg	gcccgagctt	3660
aagactggcc	gtcgtthttac	aacacagaaa	gagttttag	aaacgcaaaa	aggccatccg	3720
tcaggggcct	tctgcttagt	ttgatgctg	gcagttccct	actctgcct	tcgcttccct	3780
cgctcaactga	ctcctgcgc	tcggctgctc	ggctgcggcg	agcggatca	gctcaactcaa	3840
agggcggaat	acggttatcc	acagaatcag	gggataacgc	aggaaagaac	atgtgagcaa	3900
aaggccagca	aaaggccagg	aaccgtaaaa	agccgcgctt	gctggcgttt	ttccataggc	3960
tccgcccccc	tgacgagcat	cacaaaaatc	gacgctcaag	tcagaggtgg	cgaaaccgca	4020
caggactata	aagataccag	gcgtttcccc	ctggaagctc	cctcgtgcgc	tctcctgttc	4080
cgaccctgcc	gcttaccgga	tacctgtccg	cctthtcccc	ttcgggaagc	gtggcgcttt	4140
ctcatagctc	acgctgtagg	tatctcagtt	cggtgtaggt	cgttcgtccc	aagctgggct	4200
gtgtgcacga	acccccggtt	cagcccagcc	gctgcgcctt	atccggtaac	tatcgtcttg	4260
agtccaacc	ggtaagacac	gacttatcgc	cactggcagc	agccactggt	aacaggatta	4320
gcagagcgag	gtatgtaggc	ggtgctacag	agttcttgaa	gtggtgggct	aactacggct	4380
acactagaag	aacagtattt	ggtatctgcg	ctctgctgaa	gccagttacc	ttcgaaaaaa	4440
gagttgtag	ctcttgatcc	ggcaaaaaa	ccaccgctgg	tagcgggtgt	thttttgttt	4500
gcaagcagca	gattacgcgc	agaaaaaaag	gatctcaaga	agatcctttg	atctthtcta	4560
cgggtctga	cgctcagtg	aacgacgcgc	gcgtaactca	cgtaaggga	thttggtcat	4620
gagcttgcgc	cgteccgta	agtcagcgt	atgctctgct	thtagaaaa	ctcatcgagc	4680
atcaaatgaa	actgcaattt	atctatatca	ggattatcaa	taccatattt	ttgaaaaagc	4740

-continued

```

cgtttctgta atgaaggaga aaactcaccg aggcagttcc ataggatggc aagatcctgg 4800
tatcgggtctg cgattccgac tcgtccaaca tcaatacaac ctattaatth cccctcgtca 4860
aaaataaggt tatcaagtga gaaatcacca tgagtgcga ctgaatccgg tgagaatggc 4920
aaaagttht gcatttcttt ccagacttgt tcaacaggcc agccattacg ctctcatca 4980
aatcactcg catcaaccaa accggtattc attcgtgatt gcgcctgagc gaggcgaaat 5040
acgcgatcgc tgttaaaagg acaattacaa acaggaatcg agtgcaaccg gcgcaggaac 5100
actgccagcg catcaacaat attttcacct gaatcaggat attcttctaa tacctggaac 5160
gctgtttttc cggggatcgc agtggtgagt aaccatgcat catcaggagt acggataaaa 5220
tgcttgatgg tcggaagtgg cataaattcc gtcagccagt ttagtctgac catctcatct 5280
gtaacatcat tggcaacgct acctttgcca tgtttcagaa acaactctgg cgcacggggc 5340
ttcccataca agcgatagat tgtcgcacct gattgcccga cattatcgcg agcccattta 5400
taccatata aatcagcadc catggtggaa tttaatcgcg gcctcagcgt tcccgttga 5460
atatggctca tattcttctt tttcaatat tattgaagca tttatcaggg ttattgtctc 5520
atgagcggat acatatttga atgtatttag aaaaataaac aaataggggt cagtgttaca 5580
accaattaac caattctgaa cattatcgcg agcccattta tacctgaata tggctcataa 5640
cacccttgt ttgcctggcg gcagtagcgc ggtggtccca cctgacccca tgccgaactc 5700
agaagtgaaa cgccgtagcg ccgatggtag tgtggggact ccccatgcga gagtagggaa 5760
ctgccaggca tcaataaaaa cgaaggctc agtcgaaaga ctgggccttt cgccgggct 5820
aattaggggg tgcgccctt attcactct atagtgaagt tcctattctc tagaaagtat 5880
aggaacttct gaagtggggc ctgcagg 5907

```

```

<210> SEQ ID NO 24
<211> LENGTH: 222
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        polynucleotide

```

```

<400> SEQUENCE: 24

```

```

gcttgtagca attgctaata aaaactgcga tcgctgctga aatgagctgg aattttgtcc 60
ctctcagctc aaaaagtatc aatgattact taatgtttgt tctgcgaaa cttcttcgag 120
aacatgcatg atttcaaaaa agttgtagtt tctgttacca attgcgaatc gagaactgcc 180
taatctgccg agtatgcgat cctttagcag gaggaaaacc at 222

```

```

<210> SEQ ID NO 25
<211> LENGTH: 1777
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        polynucleotide

```

```

<400> SEQUENCE: 25

```

```

gaattctata agtaggaggt aaaaacatgc aagaactggc cctgagaagc gagctggact 60
tcaatagcga aacctataaa gatgcgtata gccgtattaa cgccattgtg atcgaaggcg 120
agcaagaagc ataccataaac tacctggaca tggcgcaact gctgccggag gacgaggctg 180
agctgattcg tttgagcaag atggagaacc gtcacaaaaa gggttttcaa gcgtgcggca 240
agaacctcaa tgtgactccg gatatggatt atgcacagca gttctttgcg gagctgcacg 300

```

-continued

```

gcaatthtca gaaggctaaa gccgagggtta agattgttac ctgcctgtct atccaaagcc 360
tgatcatcga ggcgtttgcg attgcagcct acaacattta cattccagtg gctgatccgt 420
ttgcacgtaa aatcaccgag ggtgtcgtca aggatgagta taccacactg aatttcggcg 480
aagtttggtt gaaggaacat tttgaagcaa gcaaggcggg gttggaggac gccaaacaag 540
agaacttaac gctggtctgg cagatgttga accaggtcga aaaggatgcc gaagtgtctg 600
gtatggagaa agaggctctg gtggaggact ttatgattag ctatggtgag gcaactgagca 660
acatcggtt ttctacgaga gaaatcatga agatgagcgc gtacggtctg cgtgcagcat 720
aactcgagta taagtaggag ataaaaacat gttcggcttg attggccacc tgactagcct 780
ggagcacgcg cacagcgtgg cggatcgtt tggctacggc ccgtacgcaa cccagggttt 840
agacctgtg tgtagcgcac cggccagctt tgttgagcac tttcatgtca cgagcattac 900
gggcaaacg attgagggtta aatacattga gagcgcgttt ttgccggaga tgttgattaa 960
acgtcgtatc aaagcagcga tccgtaagat tctgaacgcg atggcatttg cgcagaagaa 1020
caatttgaac attaccgcgc tgggtggctt cagcagcatt atctttgagg agtttaactc 1080
gaaggagaat cgtcaggttc gcaatgtgag cttggagttt gaccgcttca ccaccgtaa 1140
caccatact gcttacatta tctgccgtca agtcgaacag gcgagcgcga aactgggtat 1200
cgacctgtcc caagcgaccg tggcgatttg cggtgccacg ggtgatattg gcagcgcagt 1260
ttgtcgtctg ctggatcgcg aaaccgacac ccaagagctg ttcctgattg cgcgcaataa 1320
ggaacgcttg caacgtctgc aagatgaact gggtcgcccg aagatcatgg gcctggaaga 1380
ggcactgccc gaagcagaca ttattgtgtg ggttgctccc atgccgaagg gcctggagat 1440
taatgcccga accctgaaga agccgtgtct gatcattgac ggtggctacc cgaagaatct 1500
ggacacgaaa atcaagcacc cggacgtgca cttttgaag ggtggatttg tagagcattc 1560
gttgacatt gattgaaaa tcatggaaac cgtgaacatg gacgttccga gccgtcaaat 1620
gtttgcgtgc ttcgagagg cgatcttctt ggagtccgag caatggcaca cgaacttctc 1680
gtggggtcgc aatcaaatca cggtgacgaa gatggaacag attggtgagg cgagcgtgaa 1740
gcatggtctg caaccgctgc tgcctggta agaattc 1777

```

<210> SEQ ID NO 26

<211> LENGTH: 1023

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 26

```

atgttcggct tgattggcca cctgactagc ctggagcacg cgcacagcgt ggcggtgccc 60
tttggtacag gcccgtaacg aaccagggt ttagaactgt ggtgtagcgc accgccacag 120
tttgtgagc actttcatgt cagcagcatt acgggcaaaa cgattgaggg taaatacatt 180
gagagcgcgt ttttcccga gatgttgatt aaacgtcgt tcaaagcagc gatccgtaag 240
attctgaacg cgatggcatt tgcgcagaag aacaatttga acattaccgc gctgggtggc 300
ttcagcagca ttatctttga ggagttaaat ctgaaggaga atcgtcaggt tcgcaatgtg 360
agcttgaggt ttgaccgctt caccaccggt aacaccata ctgcttacat tatctgcctg 420
caagtccaac aggcgagcgc gaaactgggt atcgacctgt cccaagcagc cgtggcgatt 480
tgcggtgccg cgggtgatat tggcagcgcg gtttgcctg ggtggatcg caaacccgac 540

```

-continued

```

acccaagagc tgttcctgat tgcgcgcaat aaggaacgct tgcaacgtct gcaagatgaa 600
ctgggtcgcg gcaagatcat gggcctggaa gaggcactgc cggaagcaga cattattgtg 660
tggggtgcct ccatgccgaa gggcgtggag attaatgcgg aaaccctgaa gaagccgtgt 720
ctgatcattg acggtggcta cccgaagaat ctggacacga aatcaagca tccggacgtg 780
cacattttga aggggtggtat tgtagagcat tcgttggaca ttgattggaa aatcatggaa 840
accgtgaaca tggacgttcc gagccgctcaa atgtttgcgt gcttcgcaga ggcgatcttg 900
ctggagttcg agcaatggca cacgaacttc tcgtggggtc gcaatcaaat cacggtgacg 960
aagatggaac agattggtga ggcgagcgtg aagcatggtc tgcaaccgct gctgtcctgg 1020
taa 1023

```

<210> SEQ ID NO 27

<211> LENGTH: 340

<212> TYPE: PRT

<213> ORGANISM: Cyanothecce sp.

<400> SEQUENCE: 27

```

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu His Ala His Ser
 1           5           10          15
Val Ala Asp Ala Phe Gly Tyr Gly Pro Tyr Ala Thr Gln Gly Leu Asp
 20          25          30
Leu Trp Cys Ser Ala Pro Pro Gln Phe Val Glu His Phe His Val Thr
 35          40          45
Ser Ile Thr Gly Gln Thr Ile Glu Gly Lys Tyr Ile Glu Ser Ala Phe
 50          55          60
Leu Pro Glu Met Leu Ile Lys Arg Arg Ile Lys Ala Ala Ile Arg Lys
 65          70          75          80
Ile Leu Asn Ala Met Ala Phe Ala Gln Lys Asn Asn Leu Asn Ile Thr
 85          90          95
Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Glu Phe Asn Leu Lys
100          105          110
Glu Asn Arg Gln Val Arg Asn Val Ser Leu Glu Phe Asp Arg Phe Thr
115          120          125
Thr Gly Asn Thr His Thr Ala Tyr Ile Ile Cys Arg Gln Val Glu Gln
130          135          140
Ala Ser Ala Lys Leu Gly Ile Asp Leu Ser Gln Ala Thr Val Ala Ile
145          150          155          160
Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Asp
165          170          175
Arg Lys Thr Asp Thr Gln Glu Leu Phe Leu Ile Ala Arg Asn Lys Glu
180          185          190
Arg Leu Gln Arg Leu Gln Asp Glu Leu Gly Arg Gly Lys Ile Met Gly
195          200          205
Leu Glu Glu Ala Leu Pro Glu Ala Asp Ile Ile Val Trp Val Ala Ser
210          215          220
Met Pro Lys Gly Val Glu Ile Asn Ala Glu Thr Leu Lys Lys Pro Cys
225          230          235          240
Leu Ile Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Thr Lys Ile Lys
245          250          255
His Pro Asp Val His Ile Leu Lys Gly Gly Ile Val Glu His Ser Leu
260          265          270
Asp Ile Asp Trp Lys Ile Met Glu Thr Val Asn Met Asp Val Pro Ser

```

-continued

275	280	285	
Arg Gln Met Phe Ala Cys	Phe Ala Glu Ala Ile	Leu Leu Glu Phe Glu	
290	295	300	
Gln Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Val Thr			
305	310	315	320
Lys Met Glu Gln Ile Gly Glu Ala Ser Val Lys His Gly Leu Gln Pro			
	325	330	335
Leu Leu Ser Trp			
	340		

<210> SEQ ID NO 28
 <211> LENGTH: 696
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 28

atgcaagaac tggcctgag aagcgagctg gacttcaata gcgaaaccta taaagatgcg	60
tatagccgta ttaacgccat tgtgatcgaa ggcgagcaag aagcatacca aaactactg	120
gacatggcgc aactgctgcc ggaggacgag gctgagctga ttcgtttgag caagatggag	180
aaccgtcaca aaaagggttt tcaagcgtgc ggcaagaacc tcaatgtgac tccggatatg	240
gattatgcac agcagttctt tgcggagctg cacggcaatt ttcagaaggc taaagccgag	300
ggtaagattg ttacctgcct gctcatocaa agcctgatca tcgaggcgtt tgcgattgca	360
gcctacaaca tttacattcc agtggctgat ccgtttgac gtaaaatcac cgagggtgtc	420
gtcaaggatg agtataccca cctgaatttc ggcaagttt ggttgaagga acattttgaa	480
gcaagcaagg cggagttgga ggacgccaac aaagagaact taccgctggt ctggcagatg	540
ttgaaccagg tcgaaaagga tgcccgaagt ctgggtatgg agaaagaggc tctggtggag	600
gactttatga ttagctatgg tgaggcactg agcaacatcg gcttttctac gagagaaatc	660
atgaagatga ggcgctacgg tctgcgtgca gcataa	696

<210> SEQ ID NO 29
 <211> LENGTH: 231
 <212> TYPE: PRT
 <213> ORGANISM: Cyanothecce sp.

<400> SEQUENCE: 29

Met Gln Glu Leu Ala Leu Arg Ser Glu Leu Asp Phe Asn Ser Glu Thr	
1	15
Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu	
20	30
Gln Glu Ala Tyr Gln Asn Tyr Leu Asp Met Ala Gln Leu Leu Pro Glu	
35	45
Asp Glu Ala Glu Leu Ile Arg Leu Ser Lys Met Glu Asn Arg His Lys	
50	60
Lys Gly Phe Gln Ala Cys Gly Lys Asn Leu Asn Val Thr Pro Asp Met	
65	80
Asp Tyr Ala Gln Gln Phe Phe Ala Glu Leu His Gly Asn Phe Gln Lys	
85	95
Ala Lys Ala Glu Gly Lys Ile Val Thr Cys Leu Leu Ile Gln Ser Leu	
100	110
Ile Ile Glu Ala Phe Ala Ile Ala Ala Tyr Asn Ile Tyr Ile Pro Val	

-continued

115	120	125
Ala Asp Pro Phe Ala Arg Lys Ile Thr Glu Gly Val Val Lys Asp Glu 130 135 140		
Tyr Thr His Leu Asn Phe Gly Glu Val Trp Leu Lys Glu His Phe Glu 145 150 155 160		
Ala Ser Lys Ala Glu Leu Glu Asp Ala Asn Lys Glu Asn Leu Pro Leu 165 170 175		
Val Trp Gln Met Leu Asn Gln Val Glu Lys Asp Ala Glu Val Leu Gly 180 185 190		
Met Glu Lys Glu Ala Leu Val Glu Asp Phe Met Ile Ser Tyr Gly Glu 195 200 205		
Ala Leu Ser Asn Ile Gly Phe Ser Thr Arg Glu Ile Met Lys Met Ser 210 215 220		
Ala Tyr Gly Leu Arg Ala Ala 225 230		

<210> SEQ ID NO 30

<211> LENGTH: 1023

<212> TYPE: DNA

<213> ORGANISM: Cyanothecce sp.

<400> SEQUENCE: 30

```

atggtttggtt taattgggtca tttacaagt ttagaacacg cccactccgt tgctgatgcc      60
tttgctatg gccatacgc cactcagga cttgattgt ggtgttctgc tccaccccaa      120
ttcgtcgagc attttcatgt tactagcatc acaggacaaa ccatogaagg aaagtatata      180
gaatccgctt tcttaccaga aatgctgata aagcgacgga ttaaagcagc aattcgcaaa      240
atactgaatg cgatggcctt tgctcagaaa aataacctta acatcacagc attagggggc      300
ttttcttoga ttatttttga agaatttaat ctcaaagaga atagacaagt tcgtaatgtc      360
tctttagagt ttgatcgctt caccaccgga aacaccata ctgcttatat catttgctgt      420
caagttgaac aggcacccgc taaactaggg attgacttat cccaagcaac ggttgctatt      480
tgccggggcaa ccggagatat tggcagtgca gtgtgtcgtt ggtagatag aaaaaccgat      540
accaggaac tattcttaat tgctcgcaat aaagaacgat tacaacgact gcaagatgag      600
ttgggacggg gtaaaattat gggattggag gaggctttac ccgaagcaga tattatcgtt      660
tgggtggcga gtatgcccaa aggagtggaa attaatgccg aaactctcaa aaaaccctgt      720
ttaattatcg atggtgggta tctaagaat ttagacacaa aaattaaaca tctgatgtc      780
catatcctga aagggggaat tgtagaacat tctctagata ttgactggaa gattatggaa      840
actgtcaata tggatgttcc ttctcgtcaa atgtttgctt gttttgccga agccatttta      900
ttagagtttg aacaatggca cactaatttt tcttggggac gcaatcaaat tacagtgact      960
aaaatggaac aaataggaga agcttctgtc aaacatgggt tacaaccggt gttgagttgg      1020
taa                                                                                   1023

```

<210> SEQ ID NO 31

<211> LENGTH: 696

<212> TYPE: DNA

<213> ORGANISM: Cyanothecce sp.

<400> SEQUENCE: 31

```

atgcaagagc ttgctttacg ctcagagctt gattttaaca gcgaaaccta taaagatgct      60
tacagtcgca tcaatgctat tgtcattgaa ggggaacaag aagcctatca aaattatctt      120

```

-continued

```

gatatggcgc aacttctccc agaagacgag gctgagttaa ttcgtctctc caagatggaa 180
aaccgtcaca aaaaggcctt tcaagcctgt ggcaagaatt tgaatgtgac cccagatatg 240
gactacgctc aacaattttt tctgtaactt catggcaact tccaaaaggc aaaagccgaa 300
ggcaaaatg tcaacttgctt attaattcaa tctttgatca tcgaagcctt tgcgatcgcc 360
gcttataata tttatattcc tgtggcagat ccctttgctc gtaaaatcac cgaaggggta 420
gttaaggatg aatataccca cctcaatttt ggggaagtct ggttaaaaga gcattttgaa 480
gcctctaaag cagaattaga agacgcaaat aaagaaaatt taccocctgt ttggcaaatg 540
ctcaaccaag ttgaaaaga tgccgaagtg ttagggatgg agaagaagc cttagtggaa 600
gatttcatga ttagttagg agaagcttta agtaaatattg gtttctctac ccgtgagatc 660
atgaaaatgt ctgcttacgg gctacgggct gcttaa 696

```

<210> SEQ ID NO 32

<211> LENGTH: 339

<212> TYPE: PRT

<213> ORGANISM: *Trichodesmium erythraeum*

<400> SEQUENCE: 32

```

Met Phe Gly Leu Ile Gly His Leu Thr Asn Leu Glu His Ala Gln Ser
 1             5             10            15
Val Ala Arg Asp Leu Gly Tyr Pro Glu Tyr Ala Asp Gln Gly Leu Asp
          20             25            30
Phe Trp Cys Ser Ala Pro Pro Gln Ile Val Asp Thr Ile Lys Val Thr
          35             40            45
Ser Ile Thr Gly Gln Lys Ile Glu Gly Lys Tyr Val Glu Ser Cys Phe
          50             55            60
Leu Pro Glu Met Leu Ala Ser Ser Arg Ile Lys Ala Ala Thr Arg Lys
 65             70            75            80
Ile Met Asn Ala Met Ala His Ala Gln Lys His Gly Ile Asp Ile Thr
          85             90            95
Ala Leu Gly Gly Phe Ser Ser Ile Val Phe Glu Asn Phe Asn Leu Gln
          100            105           110
Lys Phe Lys Gln Ile Arg Asn Ile Thr Leu Glu Phe Lys Arg Phe Thr
          115           120           125
Thr Gly Asn Thr His Thr Ala Tyr Ile Val Cys Gln Gln Val Glu Gln
          130           135           140
Gly Ala Gln Lys Leu Gly Ile Asp Leu Ser Lys Ala Thr Val Ala Val
 145           150           155           160
Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Asn
          165           170           175
Thr Lys Thr Glu Val Glu Glu Leu Leu Ile Ala Arg Lys Gln Glu
          180           185           190
Arg Leu Asn Ala Leu Gln Lys Glu Leu Lys Arg Gly Lys Ile Leu Glu
          195           200           205
Leu Asn Ser Ala Leu Pro Met Ala Asp Ile Ile Val Trp Val Ala Ser
 210           215           220
Ile Pro Glu Ala Leu Glu Ile Asn Pro Asn Val Leu Lys Lys Pro Cys
 225           230           235           240
Leu Leu Ile Asp Gly Gly Tyr Pro Lys Asn Met Ala Thr Lys Val Gln
          245           250           255
Gln Glu Gly Ile Tyr Val Leu Asn Gly Gly Ile Val Glu His Ser Leu
          260           265           270

```

-continued

Asp Ile Asp Trp Lys Ile Met Lys Ile Val Asn Met Glu Val Pro Gly
 275 280 285

Arg Gln Leu Phe Ala Cys Phe Ala Glu Ser Met Leu Leu Glu Phe Glu
 290 295 300

Lys Leu Tyr Thr Asn Phe Ser Trp Gly Arg Asn Leu Ile Thr Val Glu
 305 310 315 320

Lys Met Glu Leu Ile Gly Lys Leu Ser Val Lys His Gly Phe Lys Pro
 325 330 335

Leu Met Leu

<210> SEQ ID NO 33

<211> LENGTH: 341

<212> TYPE: PRT

<213> ORGANISM: *Acaryochloris marina*

<400> SEQUENCE: 33

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Gln His Ala Gln Ser
 1 5 10 15

Val Ala Arg Glu Leu Gly Tyr Pro Glu Tyr Ala Asp Gln Gly Leu Asp
 20 25 30

Phe Trp Cys Met Ala Pro Pro Gln Ile Val Asp Asp Ile Thr Val Thr
 35 40 45

Ser Ile Thr Gly Gln Thr Ile Tyr Gly Lys Tyr Val Glu Ser Cys Phe
 50 55 60

Leu Pro Glu Met Leu Ala Ser Gln Arg Ile Lys Ala Ala Thr Arg Lys
 65 70 75 80

Ile Val Asn Ala Met Ala His Ala Gln Lys Asn Gly Ile Asn Ile Thr
 85 90 95

Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Asn Phe Asn Leu Gln
 100 105 110

Arg Ile Thr Arg Ile Arg Asn Ile Gln Leu Asp Leu Gln Arg Phe Thr
 115 120 125

Thr Gly Asn Thr His Thr Ala Tyr Ile Ile Cys Arg Gln Val Glu Gln
 130 135 140

Gly Ala Gln Lys Leu Gly Ile Asp Leu Asn Lys Ala Thr Val Ala Val
 145 150 155 160

Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Asp
 165 170 175

Ala Arg Thr Asp Thr Ala Glu Leu Leu Leu Val Ala Arg His Gln Gly
 180 185 190

Arg Leu Glu Thr Leu Gln Ser Glu Leu Gly Arg Gly Lys Ile Met Ser
 195 200 205

Ile Glu Glu Ala Leu Pro Gln Ala Asp Ile Val Val Trp Val Ala Ser
 210 215 220

Met Pro Lys Gly Ile Glu Ile Asp Ala Glu Asn Leu Lys His Pro Cys
 225 230 235 240

Leu Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Thr Lys Ile Gln
 245 250 255

His Pro Asp Val His Ile Leu Asn Gly Gly Ile Val Glu His Ser Leu
 260 265 270

Asp Ile Asp Trp Lys Ile Met His Ile Val Asn Met Asn Ile Pro Asn
 275 280 285

Arg Gln Leu Phe Ala Cys Phe Ala Glu Ser Met Leu Leu Glu Phe Glu
 290 295 300

-continued

Gln Leu His Thr Asn Phe Ser Trp Gly Arg Asn Glu Ile Thr Val Ala
305 310 315 320

Lys Met Glu Lys Ile Gly Glu Ile Ser Leu Lys His Gly Phe Lys Pro
325 330 335

Leu Ala Leu Ala Val
340

<210> SEQ ID NO 34

<211> LENGTH: 340

<212> TYPE: PRT

<213> ORGANISM: Cyanothecce sp.

<400> SEQUENCE: 34

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu His Ala Gln Gln
1 5 10 15

Val Ala Glu Ala Leu Gly Tyr Pro Glu Tyr Ala Asn Gln Gly Leu Asp
20 25 30

Phe Trp Cys Ala Ala Pro Pro Gln Ile Val Asp His Phe His Val Thr
35 40 45

Ser Val Thr Gly Gln Ile Ile Glu Gly Lys Tyr Val Glu Ser Cys Phe
50 55 60

Leu Pro Glu Met Leu Val Asn Arg Arg Ile Lys Ala Ala Ile Arg Lys
65 70 75 80

Ile Leu Asn Ala Met Ala Leu Ala Gln Lys Ala Asp Leu Asn Ile Thr
85 90 95

Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Glu Phe Asn Leu Lys
100 105 110

Glu Asn Lys Gln Val Arg Asn Val Glu Leu Glu Phe Glu Arg Phe Thr
115 120 125

Thr Gly Asn Thr His Thr Ala Tyr Ile Ile Cys Arg Gln Leu Glu Gln
130 135 140

Val Ser Ala Gln Leu Gly Leu Asp Leu Ser Gln Ala Thr Val Ala Val
145 150 155 160

Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Asp
165 170 175

Gln Lys Thr Asp Val Ala Glu Leu Leu Leu Ile Ala Arg Asn Gln Glu
180 185 190

Arg Leu Gln Gly Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Met Glu
195 200 205

Leu Glu Glu Ala Leu Pro Gln Ala Asp Ile Ile Val Trp Val Ala Ser
210 215 220

Met Pro Lys Gly Val Glu Ile Asn Pro Glu Thr Leu Lys Lys Pro Cys
225 230 235 240

Leu Ile Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Thr Gln Val Gln
245 250 255

His Pro Asp Val Tyr Val Leu Lys Gly Gly Ile Val Glu His Ser Leu
260 265 270

Asp Ile Asp Trp Lys Ile Met Glu Ile Val Ser Met Asp Ile Pro Ser
275 280 285

Arg Gln Met Phe Ala Cys Phe Ala Glu Gly Ile Leu Leu Glu Phe Glu
290 295 300

Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Ser Val Pro
305 310 315 320

Lys Met Glu Gln Ile Gly Glu Ala Ser Leu Lys His Gly Phe Arg Pro
325 330 335

-continued

Leu Leu Ser Trp
340

<210> SEQ ID NO 35

<211> LENGTH: 340

<212> TYPE: PRT

<213> ORGANISM: Cyanothecce sp.

<400> SEQUENCE: 35

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu His Ala Gln Gln
1 5 10 15

Val Ala Glu Ala Leu Gly Tyr Pro Glu Tyr Ala Asn Gln Gly Leu Asp
20 25 30

Phe Trp Cys Ala Ala Pro Pro Gln Ile Val Asp His Phe His Val Thr
35 40 45

Ser Val Thr Gly Gln Ile Ile Glu Gly Lys Tyr Val Glu Ser Cys Phe
50 55 60

Leu Pro Glu Met Leu Val Asn Arg Arg Ile Lys Ala Ala Ile Arg Lys
65 70 75 80

Ile Leu Asn Ala Met Ala Leu Ala Gln Lys Ala Asp Leu Asn Ile Thr
85 90 95

Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Glu Phe Asn Leu Lys
100 105 110

Glu Asn Lys Gln Val Arg Asn Val Glu Leu Glu Phe Glu Arg Phe Thr
115 120 125

Thr Gly Asn Thr His Thr Ala Tyr Ile Ile Cys Arg Gln Leu Glu Gln
130 135 140

Val Ser Ala Gln Leu Gly Leu Asp Leu Ser Gln Ala Thr Val Ala Val
145 150 155 160

Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Asp
165 170 175

Gln Lys Thr Asp Val Ala Glu Leu Leu Leu Ile Ala Arg Asn Gln Glu
180 185 190

Arg Leu Gln Gly Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Met Glu
195 200 205

Leu Glu Glu Ala Leu Pro Gln Ala Asp Ile Ile Val Trp Val Ala Ser
210 215 220

Met Pro Lys Gly Val Glu Ile Asn Pro Glu Thr Leu Lys Lys Pro Cys
225 230 235 240

Leu Ile Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Thr Gln Val Gln
245 250 255

His Pro Asp Val Tyr Val Leu Lys Gly Gly Ile Val Glu His Ser Leu
260 265 270

Asp Ile Asp Trp Lys Ile Met Glu Ile Val Ser Met Asp Ile Pro Ser
275 280 285

Arg Gln Met Phe Ala Cys Phe Ala Glu Gly Ile Leu Leu Glu Phe Glu
290 295 300

Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Ser Val Pro
305 310 315 320

Lys Met Glu Gln Ile Gly Glu Ala Ser Leu Lys His Gly Phe Arg Pro
325 330 335

Leu Leu Ser Trp
340

-continued

```

<210> SEQ ID NO 36
<211> LENGTH: 340
<212> TYPE: PRT
<213> ORGANISM: Microcoleus chthonoplastes

<400> SEQUENCE: 36

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Gln His Ala Gln Ala
 1           5           10           15
Val Ala Arg Asp Leu Gly Tyr Pro Glu Tyr Ala Asp Gln Gly Leu Asp
          20           25           30
Phe Trp Cys Ser Ala Pro Pro Gln Ile Val Asp Thr Ile Lys Val Thr
          35           40           45
Ser Leu Thr Gly Glu Thr Ile Glu Gly Arg Tyr Val Glu Ser Cys Phe
 50           55           60
Leu Pro Glu Met Leu Ala Thr Arg Arg Ile Lys Ala Ala Ile Arg Lys
 65           70           75           80
Val Leu Asn Ala Met Ala His Ala Gln Lys Asn Gly Ile Glu Ile Thr
          85           90           95
Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Glu Phe His Leu Tyr
          100          105          110
Glu Lys Ser Gln Val Arg Asn Ile Lys Leu Glu Phe Glu Arg Phe Thr
          115          120          125
Thr Gly Asn Thr His Thr Ala Tyr Ile Ile Ser Ala Gln Val Glu Gln
          130          135          140
Gly Ala Gln Lys Leu Gly Ile Asp Leu Ser Lys Ala Thr Val Ala Val
          145          150          155          160
Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Asn
          165          170          175
Thr Arg Thr Asp Val Ala Glu Ile Leu Leu Thr Ala Arg Asn Gln Glu
          180          185          190
Arg Leu Gln Ala Leu Gln Asp Gln Leu Gly Arg Gly Lys Ile Met Gly
          195          200          205
Leu Glu Glu Ala Leu Pro Gln Ala Asp Ile Ile Val Trp Val Ala Ser
          210          215          220
Met Ser Lys Gly Ile Asp Ile Asp Ala Ser Leu Leu Lys Lys Pro Cys
          225          230          235          240
Leu Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Ala Thr Lys Leu Gln
          245          250          255
His Pro Asp Ile Tyr Val Leu Asn Gly Gly Ile Val Glu His Ser Leu
          260          265          270
Asp Ile Asp Trp Lys Ile Met Gln Ile Val Glu Met Lys Asp Pro Gly
          275          280          285
Arg Gln Leu Phe Ala Cys Phe Ala Glu Ser Met Leu Leu Glu Phe Glu
          290          295          300
Lys Trp Tyr Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Val Glu
          305          310          315          320
Lys Met Asp Lys Ile Gly Gln Val Ser Ile Lys His Gly Phe Arg Pro
          325          330          335
Leu Leu Asn Val
          340

```

```

<210> SEQ ID NO 37
<211> LENGTH: 339
<212> TYPE: PRT
<213> ORGANISM: Arthrospira maxima

```

-continued

<400> SEQUENCE: 37

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu His Ala Gln Val
 1 5 10 15
 Val Ala Arg Asp Leu Gly Tyr Ala Glu Tyr Ala Asp Gln Gly Leu Asp
 20 25 30
 Phe Trp Cys Ser Ala Pro Pro Val Ile Val Glu Asp Leu Lys Val Thr
 35 40 45
 Ser Ile Thr Gly Gln Val Ile Glu Gly Arg Tyr Val Glu Ser Cys Phe
 50 55 60
 Leu Pro Glu Met Leu Ala Thr Asn Arg Met Lys Ala Ala Thr Arg Lys
 65 70 75 80
 Ile Ile Asn Ala Met Ala His Ala Gln Lys Asn Gly Ile Asn Ile Thr
 85 90 95
 Ala Leu Gly Gly Phe Ser Ser Ile Ile Leu Glu Arg Phe Asn Leu Asp
 100 105 110
 Gln Leu Gly Arg Ile Arg Asn Ile Lys Leu Glu Phe Glu Arg Phe Thr
 115 120 125
 Thr Gly Asn Thr His Thr Ala Tyr Ile Ile Cys Arg Gln Val Glu Gln
 130 135 140
 Ala Ala Pro Lys Leu Gly Ile Asp Leu Ser Lys Ala Thr Val Ala Val
 145 150 155 160
 Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Asn
 165 170 175
 Gly Arg Leu Asp Val Ala Glu Ile Leu Leu Ile Ala Arg Asp Arg Gln
 180 185 190
 Arg Leu Gln Asn Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Met Ala
 195 200 205
 Leu Asp Glu Ala Leu Pro Gln Ala Asp Ile Val Val Trp Val Ala Ser
 210 215 220
 Met Pro Gln Gly Val Glu Ile Asp Pro Glu Val Leu Lys Lys Pro Cys
 225 230 235 240
 Leu Leu Ile Asp Gly Gly Tyr Pro Lys Asn Met Ala Thr Lys Phe Gln
 245 250 255
 Ser Pro Gly Val His Val Leu Ser Gly Gly Ile Val Glu His Ala Leu
 260 265 270
 Asp Ile Asp Trp Lys Ile Met Lys Ile Val Asn Met Asn Val Pro Gly
 275 280 285
 Arg Gln Leu Phe Ala Cys Phe Ala Glu Ser Met Leu Leu Glu Phe Glu
 290 295 300
 Ser Ile Tyr Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Leu Asp
 305 310 315 320
 Lys Met Asp Met Ile Gly Arg Met Ser Ile Lys His Gly Phe Lys Pro
 325 330 335
 Leu Met Leu

<210> SEQ ID NO 38

<211> LENGTH: 340

<212> TYPE: PRT

<213> ORGANISM: Synechocystis sp.

<400> SEQUENCE: 38

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu His Ala Gln Ala
 1 5 10 15
 Val Ala Glu Asp Leu Gly Tyr Pro Glu Tyr Ala Asn Gln Gly Leu Asp

-continued

20					25					30					
Phe	Trp	Cys	Ser	Ala	Pro	Pro	Gln	Val	Val	Asp	Asn	Phe	Gln	Val	Lys
		35					40					45			
Ser	Val	Thr	Gly	Gln	Val	Ile	Glu	Gly	Lys	Tyr	Val	Glu	Ser	Cys	Phe
	50					55					60				
Leu	Pro	Glu	Met	Leu	Thr	Gln	Arg	Arg	Ile	Lys	Ala	Ala	Ile	Arg	Lys
65					70					75					80
Ile	Leu	Asn	Ala	Met	Ala	Leu	Ala	Gln	Lys	Val	Gly	Leu	Asp	Ile	Thr
				85					90					95	
Ala	Leu	Gly	Gly	Phe	Ser	Ser	Ile	Val	Phe	Glu	Glu	Phe	Asn	Leu	Lys
			100					105						110	
Gln	Asn	Asn	Gln	Val	Arg	Asn	Val	Glu	Leu	Asp	Phe	Gln	Arg	Phe	Thr
			115				120						125		
Thr	Gly	Asn	Thr	His	Thr	Ala	Tyr	Val	Ile	Cys	Arg	Gln	Val	Glu	Ser
	130					135					140				
Gly	Ala	Lys	Gln	Leu	Gly	Ile	Asp	Leu	Ser	Gln	Ala	Thr	Val	Ala	Val
145					150					155					160
Cys	Gly	Ala	Thr	Gly	Asp	Ile	Gly	Ser	Ala	Val	Cys	Arg	Trp	Leu	Asp
				165					170						175
Ser	Lys	His	Gln	Val	Lys	Glu	Leu	Leu	Leu	Ile	Ala	Arg	Asn	Arg	Gln
			180						185					190	
Arg	Leu	Glu	Asn	Leu	Gln	Glu	Glu	Leu	Gly	Arg	Gly	Lys	Ile	Met	Asp
		195					200						205		
Leu	Glu	Thr	Ala	Leu	Pro	Gln	Ala	Asp	Ile	Ile	Val	Trp	Val	Ala	Ser
	210					215						220			
Met	Pro	Lys	Gly	Val	Glu	Ile	Ala	Gly	Glu	Met	Leu	Lys	Lys	Pro	Cys
225					230					235					240
Leu	Ile	Val	Asp	Gly	Gly	Tyr	Pro	Lys	Asn	Leu	Asp	Thr	Arg	Val	Lys
				245					250						255
Ala	Asp	Gly	Val	His	Ile	Leu	Lys	Gly	Gly	Ile	Val	Glu	His	Ser	Leu
			260					265							270
Asp	Ile	Thr	Trp	Glu	Ile	Met	Lys	Ile	Val	Glu	Met	Asp	Ile	Pro	Ser
		275					280						285		
Arg	Gln	Met	Phe	Ala	Cys	Phe	Ala	Glu	Ala	Ile	Leu	Leu	Glu	Phe	Glu
		290				295						300			
Gly	Trp	Arg	Thr	Asn	Phe	Ser	Trp	Gly	Arg	Asn	Gln	Ile	Ser	Val	Asn
305					310					315					320
Lys	Met	Glu	Ala	Ile	Gly	Glu	Ala	Ser	Val	Lys	His	Gly	Phe	Cys	Pro
				325						330					335
Leu	Val	Ala	Leu												
			340												

<210> SEQ ID NO 39

<211> LENGTH: 340

<212> TYPE: PRT

<213> ORGANISM: Cyanosethece sp.

<400> SEQUENCE: 39

Met	Phe	Gly	Leu	Ile	Gly	His	Leu	Thr	Ser	Leu	Glu	His	Ala	Gln	Ser
1				5					10					15	
Val	Ala	Asp	Ala	Leu	Gly	Tyr	Pro	Glu	Tyr	Ala	Asn	Gln	Gly	Leu	Asp
			20					25					30		
Phe	Trp	Cys	Ser	Ala	Pro	Pro	Gln	Ile	Val	Asp	His	Phe	His	Val	Thr
		35					40					45			

-continued

Ser Val Thr Gly Gln Thr Ile Glu Gly Lys Tyr Val Glu Ser Cys Phe
 50 55 60
 Leu Pro Glu Met Leu Met Asn Arg Arg Ile Lys Ala Ala Ile Arg Lys
 65 70 75 80
 Ile Leu Asn Ala Met Ala Leu Ala Gln Lys Asn Ser Ile Asn Ile Thr
 85 90 95
 Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Glu Phe Asn Leu Lys
 100 105 110
 Asp Asn Lys Gln Val Arg Asn Val Ser Leu Glu Phe Asp Arg Phe Thr
 115 120 125
 Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu Gln
 130 135 140
 Gly Ser Ala Lys Leu Gly Ile Asp Leu Ser Lys Ala Thr Ile Ala Val
 145 150 155 160
 Cys Gly Ala Thr Gly Asp Ile Gly Ser Gly Val Cys Arg Trp Leu Asp
 165 170 175
 Arg Asn Thr Asn Thr Gln Glu Leu Leu Leu Ile Ala Arg Asn Gln Glu
 180 185 190
 Arg Leu Lys Arg Leu Gln Asp Glu Leu Gly Arg Gly Lys Ile Met Gly
 195 200 205
 Leu Glu Glu Ala Leu Pro Glu Ala Asp Val Ile Val Trp Val Ala Ser
 210 215 220
 Met Pro Lys Gly Val Glu Ile Asn Pro Glu Thr Leu Lys Lys Pro Cys
 225 230 235 240
 Leu Ile Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Thr Lys Ile Lys
 245 250 255
 His Ser Asp Val His Ile Leu Lys Gly Gly Ile Val Glu His Ser Leu
 260 265 270
 Asp Ile Asp Trp Lys Ile Met Glu Ile Val Ser Met Asp Ile Pro Ser
 275 280 285
 Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Ile Leu Leu Glu Phe Glu
 290 295 300
 Gln Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Val Thr
 305 310 315 320
 Lys Met Glu Gln Ile Gly Thr Ala Ser Val Lys His Gly Phe Gln Pro
 325 330 335
 Leu Leu Asn Trp
 340

<210> SEQ ID NO 40

<211> LENGTH: 339

<212> TYPE: PRT

<213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 40

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu His Ala Lys Ser
 1 5 10 15
 Val Ala His Lys Leu Gly Tyr Pro Glu Tyr Ala Glu Gln Gly Leu Asp
 20 25 30
 Phe Trp Cys Ser Ala Pro Pro Gln Val Val Asp His Phe Lys Val Val
 35 40 45
 Ser Ala Thr Gly Gln Thr Ile Glu Gly Lys Tyr Val Glu Ser Cys Phe
 50 55 60
 Leu Pro Glu Met Leu Ala Asn Arg Arg Ile Lys Ala Ala Thr Arg Lys
 65 70 75 80

-continued

Glu Asn Arg Gln Val Arg Asn Val Ser Leu Glu Phe Asp Arg Phe Thr
 115 120 125
 Thr Gly Asn Thr His Thr Ala Tyr Ile Ile Cys Arg Gln Val Glu Gln
 130 135 140
 Ala Ser Ala Lys Leu Gly Ile Asp Leu Ser Gln Ala Thr Val Ala Ile
 145 150 155 160
 Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Asp
 165 170 175
 Arg Lys Thr Asp Thr Gln Glu Leu Phe Leu Ile Ala Arg Asn Lys Glu
 180 185 190
 Arg Leu Gln Arg Leu Gln Asp Glu Leu Gly Arg Gly Lys Ile Met Gly
 195 200 205
 Leu Glu Glu Ala Leu Pro Glu Ala Asp Ile Ile Val Trp Val Ala Ser
 210 215 220
 Met Pro Lys Gly Val Glu Ile Asn Ala Glu Thr Leu Lys Lys Pro Cys
 225 230 235 240
 Leu Ile Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Thr Lys Ile Lys
 245 250 255
 His Pro Asp Val His Ile Leu Lys Gly Gly Ile Val Glu His Ser Leu
 260 265 270
 Asp Ile Asp Trp Lys Ile Met Glu Thr Val Asn Met Asp Val Pro Ser
 275 280 285
 Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Ile Leu Leu Glu Phe Glu
 290 295 300
 Gln Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Val Thr
 305 310 315 320
 Lys Met Glu Gln Ile Gly Glu Ala Ser Val Lys His Gly Leu Gln Pro
 325 330 335
 Leu Leu Ser Trp
 340

<210> SEQ ID NO 42

<211> LENGTH: 350

<212> TYPE: PRT

<213> ORGANISM: *Gloeobacter violaceus*

<400> SEQUENCE: 42

Met Phe Gly Leu Ile Gly His Leu Thr Asn Leu Ser His Ala Gln Arg
 1 5 10 15
 Val Ala Arg Asp Leu Gly Tyr Asp Glu Tyr Ala Ser His Asp Leu Glu
 20 25 30
 Phe Trp Cys Met Ala Pro Pro Gln Ala Val Asp Glu Ile Thr Ile Thr
 35 40 45
 Ser Val Thr Gly Gln Val Ile His Gly Gln Tyr Val Glu Ser Cys Phe
 50 55 60
 Leu Pro Glu Met Leu Ala Gln Gly Arg Phe Lys Thr Ala Met Arg Lys
 65 70 75 80
 Ile Leu Asn Ala Met Ala Leu Val Gln Lys Arg Gly Ile Asp Ile Thr
 85 90 95
 Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Asn Phe Ser Leu Asp
 100 105 110
 Lys Leu Leu Asn Val Arg Asp Ile Thr Leu Asp Ile Gln Arg Phe Thr
 115 120 125
 Thr Gly Asn Thr His Thr Ala Tyr Ile Leu Cys Gln Gln Val Glu Gln

-continued

130					135					140					
Gly	Ala	Val	Arg	Tyr	Gly	Ile	Asp	Pro	Ala	Lys	Ala	Thr	Val	Ala	Val
145					150					155					160
Val	Gly	Ala	Thr	Gly	Asp	Ile	Gly	Ser	Ala	Val	Cys	Arg	Trp	Leu	Thr
				165					170					175	
Asp	Arg	Ala	Gly	Ile	His	Glu	Leu	Leu	Leu	Val	Ala	Arg	Asp	Ala	Glu
			180						185				190		
Arg	Leu	Asp	Arg	Leu	Gln	Gln	Glu	Leu	Gly	Thr	Gly	Arg	Ile	Leu	Pro
		195					200					205			
Val	Glu	Glu	Ala	Leu	Pro	Lys	Ala	Asp	Ile	Val	Val	Trp	Val	Ala	Ser
	210					215					220				
Met	Asn	Gln	Gly	Met	Ala	Ile	Asp	Pro	Ala	Gly	Leu	Arg	Thr	Pro	Cys
225					230					235					240
Leu	Leu	Ile	Asp	Gly	Gly	Tyr	Pro	Lys	Asn	Met	Ala	Gly	Thr	Leu	Gln
				245					250					255	
Arg	Pro	Gly	Ile	His	Ile	Leu	Asp	Gly	Gly	Met	Val	Glu	His	Ser	Leu
			260					265						270	
Asp	Ile	Asp	Trp	Gln	Ile	Met	Ser	Phe	Leu	Asn	Val	Pro	Asn	Pro	Ala
		275					280					285			
Arg	Gln	Phe	Phe	Ala	Cys	Phe	Ala	Glu	Ser	Met	Leu	Leu	Glu	Phe	Glu
	290				295					300					
Gly	Leu	His	Phe	Asn	Phe	Ser	Trp	Gly	Arg	Asn	His	Ile	Thr	Val	Glu
305				310					315					320	
Lys	Met	Ala	Gln	Ile	Gly	Ser	Leu	Ser	Lys	Lys	His	Gly	Phe	Arg	Pro
				325					330					335	
Leu	Leu	Glu	Pro	Ser	Gln	Arg	Ser	Gly	Glu	Leu	Val	His	Gly		
			340					345					350		

<210> SEQ ID NO 43

<211> LENGTH: 339

<212> TYPE: PRT

<213> ORGANISM: Microcystis aeruginosa

<400> SEQUENCE: 43

Met	Phe	Gly	Leu	Ile	Gly	His	Leu	Thr	Ser	Leu	Glu	His	Ala	Gln	Ser
1				5					10					15	
Val	Ala	Asp	Asp	Leu	Gly	Tyr	Pro	Glu	Tyr	Ala	Asn	Gln	Gly	Leu	Asp
			20					25					30		
Phe	Trp	Cys	Ala	Ala	Pro	Pro	Gln	Ile	Val	Asp	Asp	Phe	His	Val	Thr
		35					40					45			
Ser	Ile	Thr	Gly	Gln	Thr	Ile	Arg	Gly	Lys	Tyr	Ile	Glu	Ser	Cys	Phe
	50					55					60				
Leu	Pro	Glu	Met	Leu	Ser	Asn	Arg	Trp	Val	Lys	Ser	Ala	Ile	Arg	Lys
65					70					75					80
Val	Leu	Asn	Ala	Met	Ala	Leu	Ala	Gln	Lys	Ser	Asp	Ile	Asn	Ile	Thr
				85					90					95	
Ala	Leu	Gly	Gly	Phe	Ser	Ser	Ile	Ile	Phe	Glu	Glu	Phe	Asn	Leu	Lys
			100						105				110		
Asp	Asn	Arg	Gln	Val	Arg	Asn	Ile	Glu	Leu	Asp	Phe	Gly	Arg	Phe	Thr
		115					120					125			
Thr	Gly	Asn	Thr	His	Thr	Ala	Tyr	Val	Ile	Cys	Thr	Gln	Val	Glu	Thr
	130					135					140				
Leu	Ala	Glu	Lys	Met	Gly	Ile	Asp	Leu	Ala	Gln	Ser	Thr	Val	Val	Val
145					150						155				160

-continued

Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Asn
 165 170 175
 Glu Lys Thr Asp Thr Lys Glu Leu Ile Cys Val Ala Arg Asn Gln Glu
 180 185 190
 Arg Leu Gln Ser Leu Gln Glu Glu Leu Gly Arg Gly Lys Ile Leu Pro
 195 200 205
 Leu Glu Glu Ala Leu Pro Leu Ala Asp Ile Ile Val Trp Val Ala Ser
 210 215 220
 Leu Pro Lys Gly Val Glu Ile Asp Pro Asp Lys Leu Lys Arg Pro Cys
 225 230 235 240
 Ile Ile Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Thr Val Leu Asn
 245 250 255
 Ala Pro Asp Ile Ser Val Ile Lys Gly Gly Ile Val Glu His Ser Leu
 260 265 270
 Asp Ile Asp Trp Lys Ile Met Lys Ile Val Asn Met Asp Ile Pro Ser
 275 280 285
 Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Ile Leu Leu Glu Leu Glu
 290 295 300
 Gly Trp Gln Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Val Pro
 305 310 315 320
 Lys Met Glu Gln Ile Gly Ala Ala Ser Arg Lys His Gly Phe Gln Pro
 325 330 335
 Leu Leu Phe

<210> SEQ ID NO 44

<211> LENGTH: 340

<212> TYPE: PRT

<213> ORGANISM: *Crocospaera watsonii*

<400> SEQUENCE: 44

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu His Ala His Ser
 1 5 10 15
 Val Ala Asp Ala Phe Gly Tyr Gly Pro Tyr Ala Thr Gln Gly Leu Asp
 20 25 30
 Leu Trp Cys Ser Ala Pro Pro Gln Phe Val Glu Gln Phe His Val Thr
 35 40 45
 Ser Ile Thr Gly Gln Thr Ile Glu Gly Lys Tyr Ile Glu Ser Ala Phe
 50 55 60
 Leu Pro Glu Met Leu Met Lys Arg Arg Ile Lys Ala Ala Ile Arg Lys
 65 70 75 80
 Ile Leu Asn Ala Met Ala Phe Ala Gln Lys Asn Asp Leu Asn Ile Thr
 85 90 95
 Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Glu Phe Asn Leu Lys
 100 105 110
 Gly Asn Arg Gln Val Arg Asn Val Ser Leu Glu Phe Asp Arg Phe Thr
 115 120 125
 Thr Gly Asn Thr His Thr Ala Tyr Ile Ile Ala Arg Gln Val Glu Gln
 130 135 140
 Ala Ser Ala Lys Leu Gly Ile Asp Leu Ser Arg Ala Thr Val Ala Val
 145 150 155 160
 Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Asp
 165 170 175
 Arg Lys Thr Asp Thr Gln Glu Leu Leu Leu Ile Ala Arg Asn Gln Glu
 180 185 190

-continued

Arg Leu Gln Arg Leu Gln Asp Glu Leu Gly Arg Gly Lys Ile Met Gly
 195 200 205
 Leu Glu Glu Ala Leu Pro Glu Ala Asp Val Ile Val Trp Val Ala Ser
 210 215 220
 Met Pro Lys Gly Val Glu Ile Asn Pro Glu Thr Leu Lys Lys Pro Cys
 225 230 235 240
 Leu Ile Val Asp Gly Gly Tyr Pro Lys Asn Leu Asp Thr Lys Ile Lys
 245 250 255
 His Pro Asp Val His Ile Leu Lys Gly Gly Val Val Glu His Ser Leu
 260 265 270
 Asp Ile Asp Trp Lys Ile Met Glu Thr Val Asn Met Asp Val Pro Ser
 275 280 285
 Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Ile Leu Leu Glu Phe Glu
 290 295 300
 Gln Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Val Thr
 305 310 315 320
 Lys Met Glu Gln Ile Gly Gly Ala Ser Val Lys His Gly Leu Gln Pro
 325 330 335
 Leu Leu Ser Trp
 340

<210> SEQ ID NO 45

<211> LENGTH: 339

<212> TYPE: PRT

<213> ORGANISM: Microcystis aeruginosa

<400> SEQUENCE: 45

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu His Ala Gln Ser
 1 5 10 15
 Val Ala Asp Asp Leu Gly Tyr Pro Glu Tyr Ala Asn Gln Gly Leu Asp
 20 25 30
 Phe Trp Cys Ala Ala Pro Pro Gln Ile Val Asp Asp Phe His Val Thr
 35 40 45
 Ser Ile Thr Gly Gln Thr Ile Thr Gly Lys Tyr Ile Glu Ser Cys Phe
 50 55 60
 Leu Pro Glu Met Leu Ser Asn Arg Trp Val Lys Ser Ala Ile Arg Lys
 65 70 75 80
 Val Leu Asn Ala Met Ala Leu Ala Gln Lys Ser Asp Ile Asn Ile Thr
 85 90 95
 Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Glu Phe Asn Leu Lys
 100 105 110
 Asp Asn Arg Gln Val Arg Asn Ile Glu Leu Asp Phe Gly Arg Phe Thr
 115 120 125
 Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Thr Gln Val Gln Thr
 130 135 140
 Leu Ala Asp Lys Met Gly Ile Asp Leu Ala Gln Ser Thr Val Val Val
 145 150 155 160
 Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Asn
 165 170 175
 Glu Lys Thr Asp Thr Lys Glu Leu Ile Cys Val Ala Arg Asn Gln Glu
 180 185 190
 Arg Leu Gln Ser Leu Gln Glu Glu Leu Gly Arg Gly Lys Ile Leu Pro
 195 200 205
 Leu Glu Glu Ala Leu Pro Leu Ala Asp Ile Ile Val Trp Val Ala Ser
 210 215 220

-continued

Met Pro Lys Gly Val Glu Ile Asp Pro Asp Lys Leu Lys Arg Pro Cys
 225 230 235 240

Leu Ile Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Thr Val Leu Asn
 245 250 255

Ala Pro Asp Val Ser Val Ile Lys Gly Gly Ile Val Glu His Ser Leu
 260 265 270

Asp Ile Asp Trp Lys Ile Met Lys Ile Val Asn Met Asp Ile Pro Ser
 275 280 285

Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Ile Leu Leu Glu Leu Glu
 290 295 300

Gly Trp Gln Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Ser Val Ser
 305 310 315 320

Lys Met Glu Gln Ile Gly Ala Ala Ser Arg Lys His Gly Phe Gln Pro
 325 330 335

Leu Leu Phe

<210> SEQ ID NO 46
 <211> LENGTH: 348
 <212> TYPE: PRT
 <213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 46

Met Phe Gly Leu Ile Gly His Ser Thr Ser Leu Glu Gln Ala Arg Ser
 1 5 10 15

Lys Ala Leu Glu Leu Gly Phe Pro Glu Tyr Ala Asp Gly Asp Leu Asp
 20 25 30

Leu Trp Cys Val Ala Pro Pro Gln Leu Val Glu Asn Val Ser Ile Thr
 35 40 45

Ser Pro Thr Gly Lys Thr Ile Glu Gly Ala Tyr Ile Asp Ser Val Phe
 50 55 60

Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
 65 70 75 80

Asn Ala Met Glu Leu Ala Gln Lys Ser Gly Ile Asp Ile Thr Ala Leu
 85 90 95

Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Lys Asn
 100 105 110

Gln Gln Ile Arg Ser Thr Ala Leu Glu Trp Glu Arg Phe Thr Thr Gly
 115 120 125

Asn Thr His Thr Ala Trp Val Ile Ala Gln Gln Val Glu Thr Asn Ala
 130 135 140

Pro Ala Leu Gly Ile Asp Leu Ser Arg Ala Lys Val Ala Val Val Gly
 145 150 155 160

Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ser Gln Asn
 165 170 175

Thr Gly Val Gly Glu Leu Leu Leu Val Ala Arg Gln Pro Gln Pro Leu
 180 185 190

Leu Asp Leu Gln Ala Glu Leu Gly Ser Gly Arg Ile Leu Ser Leu Glu
 195 200 205

Glu Ala Leu Pro Glu Ala Asp Val Val Val Trp Val Ala Ser Leu Pro
 210 215 220

Gln Gly Leu Ser Ile Asp Pro Ala Ser Leu Lys Ser Pro Cys Leu Met
 225 230 235 240

Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Ser Lys Val Thr Gly Ala
 245 250 255

-continued

```

His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Thr Leu Glu Lys Met
305                               310                               315                               320

Asp Phe Ile Gly Ala Ala Ser Val Arg His Gly Phe Ser Thr Leu Asn
                               325                               330                               335

Leu Asn Pro Ser Ala Gln Ala Ala Ala Ala
                               340                               345

```

```

<210> SEQ ID NO 49
<211> LENGTH: 346
<212> TYPE: PRT
<213> ORGANISM: Synechococcus sp.

```

```

<400> SEQUENCE: 49

```

```

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Asp Ala Ala Arg Arg
1                               5                               10                               15

Lys Ala Met Glu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu Asp
20                               25                               30

Val Trp Cys Ser Ala Pro Pro Gln Leu Val Glu His Val Glu Ile Ser
35                               40                               45

Ser Pro Thr Gly Thr Thr Ile Lys Gly Ala Tyr Ile Asp Ser Cys Phe
50                               55                               60

Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
65                               70                               75                               80

Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala Leu
85                               90                               95

Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln His
100                              105                              110

Gln Thr Val Arg Ser Thr Thr Leu Glu Trp Gln Arg Phe Thr Thr Gly
115                              120                              125

Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Val Glu Asn Asn Ala
130                              135                              140

Pro Thr Leu Gly Ile Asp Leu Lys Thr Ala Lys Val Ala Val Val Gly
145                              150                              155                              160

Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ser Ala Arg
165                              170                              175

Thr Gly Val Gly Glu Leu Leu Leu Val Ala Arg Gln Gln Gln Pro Leu
180                              185                              190

Leu Asp Leu Gln Ala Gln Ile Gly Gly Gly Arg Ile Leu Thr Leu Asp
195                              200                              205

Glu Ala Leu Pro Glu Ala Asp Val Val Val Trp Val Ala Ser Met Pro
210                              215                              220

Arg Thr Leu Glu Ile Asp Gln Ala Ser Leu Arg Lys Pro Cys Leu Met
225                              230                              235                              240

Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Thr Lys Val Ala Gly Gly
245                              250                              255

Gly Ile His Val Leu Lys Gly Gly Ile Val Glu Phe Cys Lys Asp Ile
260                              265                              270

Gly Trp Thr Met Met Gln Ile Ala Glu Met Glu Lys Pro Gln Arg Gln
275                              280                              285

Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu Arg Cys
290                              295                              300

His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Thr Leu Glu Lys Met
305                               310                               315                               320

Asp Phe Ile Gly Ala Ala Ser Val Arg His Gly Phe Ser Thr Leu Asn
325                               330                               335

```

-continued

Leu Asn Pro Ser Ala Gln Ala Ala Ala Ala
 340 345

<210> SEQ ID NO 50

<211> LENGTH: 346

<212> TYPE: PRT

<213> ORGANISM: *Synechococcus* sp.

<400> SEQUENCE: 50

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Ala Ala Arg Arg
 1 5 10 15

Lys Ala Ser Glu Leu Gly Phe Asp His Ile Ala Glu Gly Asp Leu Asp
 20 25 30

Val Trp Cys Ser Ala Pro Pro Gln Leu Val Glu His Val Glu Val Thr
 35 40 45

Ser Ala Thr Gly Arg Thr Ile Gln Gly Ala Tyr Ile Asp Ser Cys Phe
 50 55 60

Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
 65 70 75 80

Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asp Ile Thr Ala Leu
 85 90 95

Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln His
 100 105 110

Gln His Val Arg Ser Thr Thr Leu Ala Trp Glu Arg Phe Thr Thr Gly
 115 120 125

Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Val Glu Asn Asn Ala
 130 135 140

Pro Ala Leu Gly Ile Asp Leu Lys Lys Ala Ser Val Ala Val Val Gly
 145 150 155 160

Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ser Ser Arg
 165 170 175

Thr Gly Val Ala Glu Leu Leu Leu Val Ala Arg Gln Gln Lys Pro Leu
 180 185 190

Glu Glu Leu Arg Glu Glu Leu Gly Gly Gly Arg Ile Leu Ser Leu Glu
 195 200 205

Asp Ala Leu Pro Glu Ala Asp Val Val Val Trp Val Ala Ser Met Pro
 210 215 220

Arg Thr Leu Glu Ile Asp Thr Ser Arg Leu Lys Thr Pro Cys Leu Met
 225 230 235 240

Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Ala Arg Val Ala Ala Lys
 245 250 255

Gly Ile His Val Leu Lys Gly Gly Ile Val Glu Phe Phe Thr Asp Ile
 260 265 270

Gly Trp Ser Met Met Glu Ile Ala Glu Met Glu Lys Pro Gln Arg Gln
 275 280 285

Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu Ser His
 290 295 300

His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Thr Leu Glu Lys Met
 305 310 315 320

Asp Phe Ile Gly Gly Ala Ser Val Arg His Gly Phe Thr Thr Leu Asn
 325 330 335

Leu Gln Gly Leu Pro Gln Ala Ala Val Ala
 340 345

-continued

```

<210> SEQ ID NO 51
<211> LENGTH: 346
<212> TYPE: PRT
<213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 51

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Asp Ala Ala Arg Arg
 1           5           10           15
Lys Ala Met Glu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu Asp
 20           25           30
Val Trp Cys Ser Ala Pro Pro Gln Leu Val Glu His Val Glu Ile Ser
 35           40           45
Ser Pro Thr Gly Thr Thr Ile Lys Gly Ala Tyr Ile Asp Ser Cys Phe
 50           55           60
Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
 65           70           75           80
Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala Leu
 85           90           95
Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln His
 100          105          110
Gln Thr Val Arg Ser Thr Thr Leu Glu Trp Gln Arg Phe Thr Thr Gly
 115          120          125
Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Val Glu Asn Asn Ala
 130          135          140
Pro Ser Leu Gly Ile Asp Leu Lys Thr Ala Lys Val Ala Val Val Gly
 145          150          155          160
Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ser Ala Arg
 165          170          175
Thr Gly Val Gly Glu Leu Leu Leu Val Ala Arg Gln Gln Gln Pro Leu
 180          185          190
Leu Asp Leu Gln Ala Gln Ile Gly Gly Gly Arg Ile Leu Thr Leu Asp
 195          200          205
Glu Ala Leu Pro Glu Ala Asp Val Val Val Trp Val Ala Ser Met Pro
 210          215          220
Arg Thr Leu Glu Ile Asp Gln Ser Ser Leu Pro Lys Pro Cys Leu Met
 225          230          235          240
Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Ser Lys Val Ala Gly Gly
 245          250          255
Gly Ile His Val Leu Lys Gly Gly Ile Val Glu Phe Cys Lys Asp Ile
 260          265          270
Gly Trp Thr Met Met Gln Ile Ala Glu Met Asp Asn Pro Gln Arg Gln
 275          280          285
Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu Arg Cys
 290          295          300
His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Thr Leu Glu Lys Met
 305          310          315          320
Asp Phe Ile Gly Ala Ala Ser Val Arg His Gly Phe Ser Thr Leu Asn
 325          330          335
Leu Asn Pro Ser Val Gln Ala Ala Ala Ala
 340          345

```

```

<210> SEQ ID NO 52
<211> LENGTH: 346
<212> TYPE: PRT
<213> ORGANISM: Synechococcus sp.

```

-continued

<400> SEQUENCE: 52

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Ala Ala Arg Arg
 1 5 10 15
 Lys Ala Ser Asp Leu Gly Phe Asp His Ile Ala Glu Gly Asp Leu Asp
 20 25 30
 Val Trp Cys Ser Ala Pro Pro Gln Leu Val Glu His Val Glu Val Thr
 35 40 45
 Ser Pro Thr Gly Lys Ser Ile Gln Gly Ala Tyr Ile Asp Ser Cys Phe
 50 55 60
 Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
 65 70 75 80
 Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asp Ile Thr Ala Leu
 85 90 95
 Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln His
 100 105 110
 Gln His Val Arg Ser Thr Thr Leu Ala Trp Glu Arg Phe Thr Thr Gly
 115 120 125
 Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Val Glu Asn Asn Ala
 130 135 140
 Pro Ser Leu Gly Ile Asp Leu Lys Lys Ala Ser Val Ala Val Val Gly
 145 150 155 160
 Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ser Ser Arg
 165 170 175
 Thr Gly Val Ala Glu Leu Leu Leu Val Ala Arg Gln Gln Lys Pro Leu
 180 185 190
 Glu Asp Leu Arg Asp Glu Leu Gly Gly Gly Arg Ile Leu Ser Leu Glu
 195 200 205
 Asp Ala Leu Pro Glu Ala Asp Val Val Val Trp Val Ala Ser Met Pro
 210 215 220
 Arg Thr Leu Glu Ile Asp Ala Ser Arg Leu Lys Thr Pro Cys Leu Met
 225 230 235 240
 Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Ala Arg Val Ala Ala Lys
 245 250 255
 Gly Ile His Val Leu Lys Gly Gly Ile Val Glu Phe Phe Thr Asp Ile
 260 265 270
 Gly Trp Ser Met Met Glu Ile Ala Glu Met Glu Lys Pro Gln Arg Gln
 275 280 285
 Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu Ser His
 290 295 300
 His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Thr Leu Glu Lys Met
 305 310 315 320
 Asp Phe Ile Gly Gly Ala Ser Val Arg His Gly Phe Ser Thr Leu Asn
 325 330 335
 Leu Gln Gly Leu Pro Gln Ala Ala Ala Ala
 340 345

<210> SEQ ID NO 53

<211> LENGTH: 346

<212> TYPE: PRT

<213> ORGANISM: *Synechococcus* sp.

<400> SEQUENCE: 53

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Asp Ala Ala Arg Arg
 1 5 10 15

-continued

Lys Ala Leu Glu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu Asp
 20 25 30
 Val Trp Cys Ser Ala Pro Pro Gln Leu Val Glu His Val Glu Ile Thr
 35 40 45
 Ser Pro Val Gly Thr Thr Ile Lys Gly Ala Tyr Ile Asp Ser Cys Phe
 50 55 60
 Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
 65 70 75 80
 Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala Leu
 85 90
 Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln His
 100 105 110
 Gln Thr Ile Arg Ser Thr Thr Leu Glu Trp Gln Arg Phe Thr Thr Gly
 115 120 125
 Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Val Glu Asn Asn Ala
 130 135 140
 Pro Ala Leu Gly Ile Asp Leu Lys Thr Ala Lys Val Ala Val Val Gly
 145 150 155 160
 Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Thr Ala Arg
 165 170 175
 Thr Gly Val Gly Glu Leu Leu Leu Val Ala Arg Gln Gln Gln Pro Leu
 180 185 190
 Leu Asp Leu Gln Gly Glu Leu Gly Gly Gly Arg Ile Leu Ser Leu Asp
 195 200 205
 Glu Ala Met Pro Glu Ala Asp Val Val Val Trp Val Ala Ser Met Pro
 210 215 220
 Arg Thr Leu Gln Ile Asp Gln Glu Ser Leu Arg Lys Pro Cys Leu Met
 225 230 235 240
 Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Ala Lys Val Ala Gly Gly
 245 250 255
 Gly Ile His Val Leu Lys Gly Gly Ile Val Glu Phe Cys Lys Asp Ile
 260 265 270
 Gly Trp Thr Met Met Gln Ile Ala Glu Met Glu Lys Pro Gln Arg Gln
 275 280 285
 Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu Arg Cys
 290 295 300
 His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Thr Leu Glu Lys Met
 305 310 315 320
 Asp Phe Ile Gly Gln Ala Ser Val Arg His Gly Phe Ser Thr Leu Asn
 325 330 335
 Leu Asn Pro Ser Leu Gln Val Ala Ala Ala
 340 345

<210> SEQ ID NO 54

<211> LENGTH: 346

<212> TYPE: PRT

<213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 54

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Ala Ala Arg Gln
 1 5 10 15
 Lys Ala Phe Glu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu Asp
 20 25 30
 Val Trp Cys Ser Ala Pro Pro Gln Leu Val Glu Thr Phe Asp Val Thr
 35 40 45

-continued

Ser Pro Thr Gly Arg Thr Ile Thr Gly Ala Tyr Ile Asp Ser Cys Phe
 50 55 60
 Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
 65 70 75 80
 Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala Leu
 85 90 95
 Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln His
 100 105 110
 Gln His Val Arg Ser Thr Thr Leu Glu Trp Glu Arg Phe Thr Thr Gly
 115 120 125
 Asn Thr His Thr Ala Trp Val Ile Ser Arg Gln Val Glu Ile Asn Ala
 130 135 140
 Pro Arg Leu Gly Ile Asp Leu Ser Lys Ala Arg Val Ala Val Val Gly
 145 150 155 160
 Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ser Gln Arg
 165 170 175
 Thr Gly Val Ala Glu Leu Leu Leu Val Ala Arg Gln Gln Gln Pro Leu
 180 185 190
 Leu Asp Leu Gln Lys Glu Leu Gly Gly Gly Arg Ile Leu Ser Leu Asp
 195 200 205
 Glu Ala Val Pro Glu Ala Asp Val Val Val Trp Val Ala Ser Met Pro
 210 215 220
 Arg Thr Leu Glu Ile Asp Ala Ala Ser Leu Arg Gln Pro Cys Leu Met
 225 230 235 240
 Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asn Ala Arg Ile Ala Gly Ser
 245 250 255
 Gly Val His Val Leu Lys Gly Gly Ile Val Glu Phe Gly Ser Asp Ile
 260 265 270
 Gly Trp Asn Met Met Glu Leu Ala Glu Met Glu Lys Pro Gln Arg Gln
 275 280 285
 Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu Ser Cys
 290 295 300
 His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Thr Leu Glu Lys Met
 305 310 315 320
 Asp Phe Ile Gly Glu Ala Ser Arg Arg His Gly Phe Ser Thr Leu Asn
 325 330 335
 Leu Ser Ala Pro Val Gln Val Ala Ala Ala
 340 345

<210> SEQ ID NO 55

<211> LENGTH: 346

<212> TYPE: PRT

<213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 55

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Ala Ala Arg Arg
 1 5 10 15
 Lys Ala Leu Glu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu Asp
 20 25 30
 Val Trp Cys Ser Ala Pro Pro Gln Leu Val Glu His Val Glu Val Thr
 35 40 45
 Ser Pro Ala Gly Ile Thr Ile Glu Gly Ala Tyr Ile Asp Ser Cys Phe
 50 55 60
 Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu

-continued

65	70	75	80
Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala Leu	85	90	95
Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln His	100	105	110
Gln Thr Val Arg Ser Thr Thr Leu Asp Trp Gln Arg Phe Thr Thr Gly	115	120	125
Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Val Glu Asn Asn Ala	130	135	140
Pro Ser Leu Gly Ile Asp Leu Lys Thr Ala Lys Val Ala Val Val Gly	145	150	155
Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Thr Ala Arg	165	170	175
Thr Asn Val Gly Glu Leu Leu Leu Val Ala Arg Gln Pro Gln Pro Leu	180	185	190
Ala Asp Leu Gln Ala Glu Leu Gly Gly Gly Arg Ile Leu Ala Leu Ser	195	200	205
Asp Ala Leu Ser Glu Ala Asp Val Val Val Trp Val Ala Ser Met Pro	210	215	220
Arg Thr Leu Glu Ile Asp Asn Asn Ser Leu Lys Lys Pro Cys Leu Met	225	230	235
Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Ser Lys Val Ala Gly Gly	245	250	255
Gly Ile His Val Leu Lys Gly Gly Ile Val Glu Phe Cys Arg Asp Ile	260	265	270
Gly Trp Ser Met Met Glu Ile Ala Glu Met Glu Lys Pro Gln Arg Gln	275	280	285
Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu Arg Cys	290	295	300
His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Thr Leu Glu Lys Met	305	310	315
Asp Phe Ile Gly Ala Ala Ser Val Arg His Gly Phe Ser Thr Leu Asn	325	330	335
Leu Lys Thr Asn Leu Gln Ala Ala Val Ala	340	345	

<210> SEQ ID NO 56

<211> LENGTH: 347

<212> TYPE: PRT

<213> ORGANISM: Prochlorococcus marinus

<400> SEQUENCE: 56

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Lys Asp Ala Arg Gln	1	5	10	15
Lys Ala Met Asp Leu Gly Tyr Asp His Ile Ala Glu Gly Asp Leu Asp	20	25	30	
Val Trp Cys Ser Ala Pro Pro Gln Leu Val Glu His Val Lys Val Val	35	40	45	
Ser Ala Ile Gly Lys Thr Ile Glu Gly Ala Tyr Ile Asp Ser Cys Phe	50	55	60	
Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu	65	70	75	80
Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala Leu	85	90	95	

-continued

Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Glu Asn
 100 105 110
 Lys Gln Val Arg Asn Thr Thr Leu Asp Trp Glu Arg Phe Thr Thr Gly
 115 120 125
 Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Ile Asn Ala
 130 135 140
 Pro Leu Leu Gly Ile Asp Leu Gln Lys Ala Arg Val Ala Val Val Gly
 145 150 155 160
 Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ser Gln Arg
 165 170 175
 Thr Gly Val Ser Glu Leu Leu Leu Val Ala Arg Gln Gln Gln Pro Leu
 180 185 190
 Lys Asp Leu Gln Lys Asp Leu Gly Gly Gly Arg Val Leu Arg Leu Glu
 195 200 205
 Glu Ala Leu Pro Glu Ala Asp Ala Val Ile Trp Val Ala Ser Leu Pro
 210 215 220
 Lys Asn Leu Gln Ile Asp Lys Ser Lys Leu Arg Lys Pro Cys Leu Met
 225 230 235 240
 Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Ser Gly Ser
 245 250 255
 Gly Ile His Val Leu Lys Gly Gly Ile Val Glu Phe Phe Glu Asp Ile
 260 265 270
 Gly Trp Asn Met Met Glu Ile Ala Glu Met Asp Val Pro Gln Arg Gln
 275 280 285
 Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu Asn Cys
 290 295 300
 His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Thr Leu Glu Lys Met
 305 310 315 320
 Asp Phe Ile Gly Glu Ala Ser Leu Arg His Gly Phe Ser Val Leu Arg
 325 330 335
 Leu Gln Pro Asn Asn Leu Gln Ala Ala Phe Ala
 340 345

<210> SEQ ID NO 57

<211> LENGTH: 345

<212> TYPE: PRT

<213> ORGANISM: Prochlorococcus marinus

<400> SEQUENCE: 57

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Arg Lys
 1 5 10 15
 Thr Ala Leu Gln Ile Gly Tyr Asp His Leu Asp Gly Asp Leu Asp Val
 20 25 30
 Trp Cys Ser Ala Pro Pro Gln Phe Leu Glu Gln Ile Glu Val Glu Ser
 35 40 45
 Leu Thr Gly Lys Lys Ile Glu Gly Ala Tyr Ile Asp Ser Cys Phe Val
 50 55 60
 Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu Asn
 65 70 75 80
 Ala Met Glu Met Ala Gln Lys Arg Gly Ile Gln Ile Ser Ala Leu Gly
 85 90 95
 Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Lys His Gln
 100 105 110
 His Val Arg Asn Thr Thr Leu Glu Trp Glu Arg Phe Thr Thr Gly Asn
 115 120 125

-continued

Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Asn Asn Ala Pro
 130 135 140
 Leu Leu Gly Ile Asp Leu Ser Lys Ala Arg Val Ala Val Val Gly Ala
 145 150 155 160
 Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ser Ala Arg Thr
 165 170 175
 Gly Val Ala Glu Leu Leu Leu Val Ala Arg Gln Gln Gln Pro Leu Ile
 180 185 190
 Asp Leu Gln Thr Glu Leu Ala Gly Gly Arg Ile Leu Ser Leu Glu Glu
 195 200 205
 Ala Leu Pro Glu Ala Asp Val Val Val Trp Val Ala Ser Met Pro Arg
 210 215 220
 Thr Leu Glu Ile Asp Met Glu Ser Leu Arg Lys Pro Cys Leu Met Ile
 225 230 235 240
 Asp Gly Gly Tyr Pro Lys Asn Leu Asp Ala Lys Phe Ala Gly Ser Gly
 245 250 255
 Val His Val Leu Lys Gly Gly Ile Val Glu Phe Cys Asn Asp Ile Ser
 260 265 270
 Trp Asp Val Gly Trp Ile Ala Glu Met Asp Lys Pro Ala Arg Gln Met
 275 280 285
 Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu Asn Cys His
 290 295 300
 Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Leu Glu Lys Met Asp
 305 310 315 320
 Phe Ile Gly Met Ala Ser Leu Arg His Gly Phe Ser Ser Leu Asn Leu
 325 330 335
 Asn His Gln Leu Gln Ala Ala Ala Ala
 340 345

<210> SEQ ID NO 58

<211> LENGTH: 346

<212> TYPE: PRT

<213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 58

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Asp Ala Ala Arg Arg
 1 5 10 15
 Lys Ala Met Glu Leu Gly Phe Asp His Ile Ala Glu Gly Asp Leu Asp
 20 25 30
 Val Trp Cys Ser Ala Pro Pro Gln Leu Val Glu His Val Gln Val Thr
 35 40 45
 Ser Pro Val Gly Thr Thr Ile Glu Gly Ala Tyr Ile Asp Ser Cys Phe
 50 55 60
 Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
 65 70 75 80
 Asn Ala Met Glu Leu Ala Gln Lys Lys Asp Ile Ala Ile Thr Ala Leu
 85 90 95
 Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln Asn
 100 105 110
 Gln Thr Val Arg Ser Thr Thr Leu Asp Trp Arg Arg Phe Thr Thr Gly
 115 120 125
 Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Val Glu Asn Asn Ala
 130 135 140
 Pro Ser Leu Gly Ile Asp Leu Ser Thr Ala Lys Val Ala Val Val Gly

-continued

145	150	155	160
Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ser Ala Arg	165	170	175
Thr Gly Val Gly Glu Leu Leu Leu Val Ala Arg Gln Gln Gln Pro Leu	180	185	190
Met Asp Leu Gln Lys Glu Leu Gly Gly Gly Arg Ile Leu Thr Leu Glu	195	200	205
Glu Ala Leu Pro Glu Ala Asp Val Val Val Trp Val Ala Ser Met Pro	210	215	220
Arg Thr Leu Gln Ile Asp Gln Asp Ser Leu Arg Ser Pro Cys Leu Met	225	230	235
Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Ala Lys Val Ala Gly Gly	245	250	255
Gly Ile His Val Leu Lys Gly Gly Ile Val Glu Phe Cys Arg Asp Ile	260	265	270
Gly Trp Thr Met Met Glu Ile Ala Glu Met Glu Lys Pro Gln Arg Gln	275	280	285
Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu Arg Cys	290	295	300
His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Thr Leu Glu Lys Met	305	310	315
Asp Phe Ile Gly Ala Ala Ser Val Arg His Gly Phe Ser Thr Leu Asn	325	330	335
Leu Gln Ser Arg Leu Gln Ala Ala Ala Ala	340	345	

<210> SEQ ID NO 59

<211> LENGTH: 346

<212> TYPE: PRT

<213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 59

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Asp Ala Ala Arg Arg	1	5	10	15
Lys Ala Met Glu Leu Gly Phe Asp His Ile Ala Glu Gly Asp Leu Asp	20	25	30	
Val Trp Cys Ser Ala Pro Pro Gln Leu Val Glu His Val Gln Val Thr	35	40	45	
Ser Pro Val Gly Thr Thr Ile Glu Gly Ala Tyr Ile Asp Ser Cys Phe	50	55	60	
Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu	65	70	75	80
Asn Ala Met Glu Leu Ala Gln Lys Lys Asp Ile Ala Ile Thr Ala Leu	85	90	95	
Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln Asn	100	105	110	
Gln Thr Val Arg Ser Thr Thr Leu Asp Trp Arg Arg Phe Thr Thr Gly	115	120	125	
Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Val Glu Asn Asn Ala	130	135	140	
Pro Ser Leu Gly Ile Asp Leu Ser Thr Ala Lys Val Ala Val Val Gly	145	150	155	160
Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ser Ala Arg	165	170	175	

-continued

Thr Gly Val Gly Glu Leu Leu Leu Val Ala Arg Gln Gln Gln Pro Leu
 180 185 190
 Met Asp Leu Gln Lys Glu Leu Gly Gly Gly Arg Ile Leu Thr Leu Glu
 195 200 205
 Glu Ala Leu Pro Glu Ala Asp Val Val Val Trp Val Ala Ser Met Pro
 210 215 220
 Arg Thr Leu Gln Ile Asp Gln Asp Ser Leu Arg Ser Pro Cys Leu Met
 225 230 235 240
 Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Ala Lys Val Ala Gly Gly
 245 250 255
 Gly Ile His Val Leu Lys Gly Gly Ile Val Glu Phe Cys Arg Asp Ile
 260 265 270
 Gly Trp Thr Met Met Glu Ile Ala Glu Met Glu Lys Pro Gln Arg Gln
 275 280 285
 Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu Arg Cys
 290 295 300
 His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Thr Leu Glu Lys Met
 305 310 315 320
 Asp Phe Ile Gly Ala Ala Ser Val Arg His Gly Phe Ser Thr Leu Asn
 325 330 335
 Leu Gln Ser Arg Leu Gln Ala Ala Ala Ala
 340 345

<210> SEQ ID NO 60

<211> LENGTH: 346

<212> TYPE: PRT

<213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 60

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Ala Ala Arg Arg
 1 5 10 15
 Lys Ala Leu Glu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu Asp
 20 25 30
 Val Trp Cys Ser Ala Pro Pro Gln Leu Val Glu His Val Glu Val Thr
 35 40 45
 Ser Pro Ala Gly Ile Thr Ile Glu Gly Ala Tyr Ile Asp Ser Cys Phe
 50 55 60
 Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
 65 70 75 80
 Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala Leu
 85 90 95
 Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln His
 100 105 110
 Gln Thr Ile Arg Ser Thr Thr Leu Asp Trp Gln Arg Phe Thr Thr Gly
 115 120 125
 Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Val Glu Asn Asn Ala
 130 135 140
 Pro Ser Leu Gly Ile Asp Leu Lys Thr Ala Lys Val Ala Val Val Gly
 145 150 155 160
 Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ser Ala Arg
 165 170 175
 Thr Asn Val Gly Glu Leu Leu Leu Val Ala Arg Gln Pro Gln Pro Leu
 180 185 190
 Ala Asp Leu Gln Ser Glu Leu Gly Gly Gly Arg Ile Leu Ala Leu Ser
 195 200 205

-continued

Asp Ala Leu Ser Glu Ala Asp Val Val Val Trp Val Ala Ser Met Pro
 210 215 220
 Arg Thr Leu Glu Ile Asp Asn Asn Ser Leu Lys Lys Pro Cys Leu Met
 225 230 235 240
 Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Ser Lys Val Ala Gly Gly
 245 250 255
 Gly Ile His Val Leu Lys Gly Gly Ile Val Glu Phe Cys Arg Asp Ile
 260 265 270
 Gly Trp Ser Met Met Glu Ile Ala Glu Met Glu Asn Pro Gln Arg Gln
 275 280 285
 Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu Arg Cys
 290 295 300
 His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Thr Leu Glu Lys Met
 305 310 315 320
 Asp Phe Ile Gly Ala Ala Ser Val Arg His Gly Phe Ser Thr Leu Asn
 325 330 335
 Leu Lys Thr Asn Leu Gln Ala Ala Ala Ala
 340 345

<210> SEQ ID NO 61

<211> LENGTH: 349

<212> TYPE: PRT

<213> ORGANISM: Cyanobium sp.

<400> SEQUENCE: 61

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Glu Ala Arg Ala
 1 5 10 15
 Lys Ala Arg Ser Leu Gly Phe Asp Glu Tyr Ala Asp Gly Asp Leu Asp
 20 25 30
 Met Trp Cys Ala Ala Pro Pro Gln Leu Val Glu Lys Val Thr Val Thr
 35 40 45
 Ser Arg Thr Gly Lys Thr Ile Glu Gly Ala Tyr Ile Asp Ser Val Phe
 50 55 60
 Val Pro Glu Met Leu Arg Arg Phe Lys Thr Ala Lys Arg Lys Val Leu
 65 70 75 80
 Lys Ala Met Glu Leu Ala Gln Arg Ser Gly Ile Asp Ile Thr Ala Leu
 85 90 95
 Gly Gly Phe Thr Ser Ile Ile Phe Glu Asp Met Asn Leu Leu Arg Glu
 100 105 110
 Glu Arg Val Ser Ala Val Gln Leu Asn Trp Glu Arg Phe Thr Thr Gly
 115 120 125
 Asn Thr His Thr Ala Trp Val Ile Cys Gln Gln Val Glu Arg Asn Ala
 130 135 140
 Ser Ser Leu Gly Ile Asp Leu Ala Ser Ala Lys Val Ala Val Val Gly
 145 150 155 160
 Ala Ser Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Gln Arg Arg
 165 170 175
 Gly Val Gly Glu Leu Leu Leu Val Ala Arg Arg Pro Gln Pro Leu Val
 180 185 190
 Asp Leu Gln Glu Ser Leu Gly Glu Gly Arg Ile Leu Asp Leu Glu Ala
 195 200 205
 Ala Leu Pro Glu Ala Asp Val Val Val Trp Val Ala Ser Leu Pro Gln
 210 215 220
 Ser Leu Gln Ile Asp Thr Ala Ser Leu Lys Arg Pro Cys Leu Met Ile

-continued

```

225                230                235                240
Asp Gly Gly Tyr Pro Lys Asn Leu Asp Ala Lys Ala Ala Ala Glu Gly
      245                250                255
Val His Val Leu Lys Gly Gly Ile Val Glu Phe Trp Gln Asp Ile Gly
      260                265                270
Trp Gln Met Met Glu Val Ala Glu Met Ala Val Pro Gln Arg Gln Met
      275                280                285
Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Asp Phe Glu Asp Leu His
      290                295                300
Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Thr Leu Ala Ala Met Asp
      305                310                315                320
Arg Ile Gly Glu Ala Ser Leu Arg His Gly Phe Glu Ala Leu Gly Leu
      325                330                335
Gln His Ala Gly Ala Val Ser Pro Ala Leu Ala Ala Ala
      340                345

```

<210> SEQ ID NO 62

<211> LENGTH: 346

<212> TYPE: PRT

<213> ORGANISM: Prochlorococcus marinus

<400> SEQUENCE: 62

```

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys Arg
 1      5      10      15
Lys Ala Leu Gly Leu Gly Tyr Asp His Ile Ala Glu Gly Asp Leu Asp
 20     25     30
Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Lys Val Val
 35     40     45
Ser Ala Ile Gly Lys Thr Ile Glu Gly Ala Tyr Ile Asp Ser Cys Phe
 50     55     60
Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
 65     70     75     80
Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Ser Ile Thr Ala Leu
 85     90     95
Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln Asn
100    105    110
Gln Gln Val Arg Asn Thr Thr Leu Asp Trp Gln Arg Phe Thr Thr Gly
115    120    125
Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Gln Asn Ala
130    135    140
Pro Arg Ile Gly Ile Asp Leu Ser Lys Ser Lys Val Ala Val Val Gly
145    150    155    160
Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ser Asn Arg
165    170    175
Thr Gly Val Ser Glu Leu Leu Leu Val Ala Arg Gln Gln Lys Pro Leu
180    185    190
Leu Glu Leu Gln Ser Gln Leu Gly Gly Gly Arg Ile Leu Ser Leu Asp
195    200    205
Asp Ala Leu Pro Glu Ala Asp Ile Val Ile Trp Val Ala Ser Met Pro
210    215    220
Lys Thr Leu Glu Ile Asp Pro Ser Lys Ile Lys Arg Pro Cys Leu Met
225    230    235    240
Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Glu Lys Phe Ser Gly Pro
245    250    255

```

-continued

Gly Ile His Val Leu Lys Gly Gly Ile Val Gln Phe Phe Lys Asp Ile
 260 265 270

Gly Trp Ser Met Met Glu Leu Ala Glu Met Glu Asn Pro Lys Arg Glu
 275 280 285

Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu Asn Cys
 290 295 300

His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Thr Leu Glu Lys Met
 305 310 315 320

Asp Phe Ile Gly Lys Ala Ser Glu Arg His Gly Phe Ser Ala Val Gly
 325 330 335

Leu Lys Ser Asn Ile Gln Thr Leu Thr Val
 340 345

<210> SEQ ID NO 63
 <211> LENGTH: 346
 <212> TYPE: PRT
 <213> ORGANISM: Prochlorococcus marinus

<400> SEQUENCE: 63

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys Arg
 1 5 10 15

Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu Asp
 20 25 30

Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val Lys
 35 40 45

Ser Ala Thr Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys Phe
 50 55 60

Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
 65 70 75 80

Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala Leu
 85 90 95

Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln His
 100 105 110

Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr Gly
 115 120 125

Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Ile Asn Ala
 130 135 140

Pro Lys Val Gly Ile Glu Leu Lys Lys Ala Thr Val Ala Val Val Gly
 145 150 155 160

Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn Lys
 165 170 175

Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro Leu
 180 185 190

Asp Asn Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Ser Leu Asp
 195 200 205

Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met Pro
 210 215 220

Lys Thr Met Glu Ile Asn Thr Asn Asn Leu Lys Gln Pro Cys Leu Met
 225 230 235 240

Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly Asn
 245 250 255

Asn Ile His Val Val Lys Gly Gly Ile Val Lys Phe Phe Asn Asp Ile
 260 265 270

Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg Glu
 275 280 285

-continued

Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys Cys
 290 295 300
 His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys Met
 305 310 315 320
 Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile Gly
 325 330 335
 Leu Asp Lys His Pro Lys Val Leu Ala Val
 340 345

<210> SEQ ID NO 64
 <211> LENGTH: 346
 <212> TYPE: PRT
 <213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 64

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Ala Ala Arg Arg
 1 5 10 15
 Lys Ala Met Glu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu Asp
 20 25 30
 Val Trp Cys Ser Ala Pro Pro Gln Leu Val Glu His Val Glu Val Thr
 35 40 45
 Ser Pro Val Gly Thr Thr Ile Glu Gly Ala Tyr Ile Asp Ser Cys Phe
 50 55 60
 Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
 65 70 75 80
 Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala Leu
 85 90 95
 Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln His
 100 105 110
 Gln Thr Val Arg Ser Thr Thr Leu Asp Trp Gln Arg Phe Thr Thr Gly
 115 120 125
 Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Val Glu Asn Asn Ala
 130 135 140
 Pro Thr Leu Gly Ile Asp Leu Ser Lys Ala Lys Val Ala Val Val Gly
 145 150 155 160
 Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Gln Ala Arg
 165 170 175
 Thr Arg Val Gly Glu Leu Leu Leu Val Ala Arg Gln Gln Gln Pro Leu
 180 185 190
 Leu Asp Leu Gln Gln Glu Leu Gly Gly Gly Arg Ile Leu Ser Leu Asp
 195 200 205
 Glu Ala Leu Pro Glu Ala Asp Val Val Val Trp Val Ala Ser Met Pro
 210 215 220
 Arg Thr Leu Glu Ile Asp Gln Asp Ser Leu Lys Lys Pro Cys Leu Met
 225 230 235 240
 Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Thr Lys Val Ala Gly Gly
 245 250 255
 Gly Ile His Val Leu Lys Gly Gly Ile Val Glu Phe Cys Arg Asp Ile
 260 265 270
 Gly Trp Ser Met Met Ala Ile Ala Glu Met Glu Arg Pro Gln Arg Gln
 275 280 285
 Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu Arg Cys
 290 295 300
 His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Thr Leu Glu Lys Met

-continued

Leu Asn Pro Lys Leu Gln Ala Ala Ile Ala
 340 345

<210> SEQ ID NO 66
 <211> LENGTH: 346
 <212> TYPE: PRT
 <213> ORGANISM: Prochlorococcus marinus

<400> SEQUENCE: 66

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys Arg
 1 5 10 15
 Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu Asp
 20 25 30
 Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val Lys
 35 40 45
 Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys Phe
 50 55 60
 Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
 65 70 75 80
 Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala Leu
 85 90 95
 Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln His
 100 105 110
 Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr Gly
 115 120 125
 Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn Ala
 130 135 140
 Pro Lys Ile Gly Ile Asp Leu Lys Ser Ala Thr Val Ala Val Val Gly
 145 150 155 160
 Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn Lys
 165 170 175
 Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro Leu
 180 185 190
 Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu Asp
 195 200 205
 Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met Pro
 210 215 220
 Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu Met
 225 230 235 240
 Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly Asn
 245 250 255
 Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp Ile
 260 265 270
 Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg Glu
 275 280 285
 Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys Cys
 290 295 300
 His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys Met
 305 310 315 320
 Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile Gly
 325 330 335
 Leu Asp Lys His Pro Lys Val Leu Ala Val
 340 345

<210> SEQ ID NO 67

-continued

<211> LENGTH: 346
 <212> TYPE: PRT
 <213> ORGANISM: *Prochlorococcus marinus*
 <400> SEQUENCE: 67

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys Arg
 1 5 10 15
 Lys Ala Ser Met Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu Asp
 20 25 30
 Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val Lys
 35 40 45
 Ser Ala Thr Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys Phe
 50 55 60
 Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
 65 70 75 80
 Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala Leu
 85 90 95
 Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln His
 100 105 110
 Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr Gly
 115 120 125
 Asn Thr His Thr Ala Trp Val Ile Cys Lys Gln Leu Glu Ile Asn Ala
 130 135 140
 Pro Arg Ile Gly Ile Asp Leu Lys Lys Ala Thr Val Ala Val Ile Gly
 145 150 155 160
 Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn Lys
 165 170 175
 Thr Gly Ile Ser Glu Leu Leu Met Val Ala Arg Gln Gln Glu Pro Leu
 180 185 190
 Ala Leu Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Thr Ser Leu Asp
 195 200 205
 Glu Ala Leu Pro Gln Ala Asp Ile Val Val Trp Val Ala Ser Met Pro
 210 215 220
 Lys Thr Ile Glu Ile Asp Thr Asp Asn Leu Lys Lys Pro Cys Leu Met
 225 230 235 240
 Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly Glu
 245 250 255
 Asn Ile Tyr Val Leu Lys Gly Gly Ile Val Glu Phe Phe Asn Asp Ile
 260 265 270
 Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg Glu
 275 280 285
 Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys Cys
 290 295 300
 His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys Met
 305 310 315 320
 Glu Phe Ile Gly Ala Ala Ser Leu Lys His Gly Phe Ser Ala Ile Gly
 325 330 335
 Leu Asp Lys Gln Pro Lys Val Leu Thr Val
 340 345

<210> SEQ ID NO 68
 <211> LENGTH: 346
 <212> TYPE: PRT
 <213> ORGANISM: *Synechococcus* sp.
 <400> SEQUENCE: 68

-continued

```

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Ala Ala Arg Arg
 1           5              10              15
Lys Ala Ser Glu Leu Gly Phe Asp His Ile Ala Glu Gly Asp Leu Asp
          20              25              30
Val Trp Cys Ser Ala Pro Pro Gln Leu Val Glu His Val Glu Val Thr
          35              40              45
Ser Ala Thr Gly Lys Thr Ile Thr Gly Ala Tyr Ile Asp Ser Cys Phe
          50              55              60
Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
 65              70              75              80
Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala Leu
          85              90              95
Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln His
          100             105             110
Gln His Val Arg Ser Thr Thr Leu Glu Trp Glu Arg Phe Thr Thr Gly
          115             120             125
Asn Thr His Thr Ala Trp Val Ile Ser Arg Gln Val Glu Asn Asn Ala
          130             135             140
Pro Leu Leu Gly Ile Asp Leu Ser Ser Ala Lys Val Ala Val Val Gly
          145             150             155             160
Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ser Gln Arg
          165             170             175
Thr Gly Val Gly Glu Leu Leu Leu Val Ala Arg Gln Gln Gln Pro Leu
          180             185             190
Leu Asp Leu Gln Gln Glu Leu Gly Gly Gly Arg Ile Leu Ser Leu Asp
          195             200             205
Glu Ala Leu Pro Glu Ala Asp Val Val Val Trp Val Ala Ser Met Pro
          210             215             220
Arg Thr Leu Glu Ile Asp Ala Ala Ser Leu Arg Lys Pro Cys Leu Met
          225             230             235             240
Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Ala Lys Val Ala Ser Ala
          245             250             255
Gly Val His Val Leu Lys Gly Gly Ile Val Glu Phe Gly Ser Asp Ile
          260             265             270
Gly Trp Ser Met Met Glu Ile Ala Glu Met Glu Lys Pro Gln Arg Gln
          275             280             285
Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Asp Phe Glu Glu Cys
          290             295             300
His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Thr Leu Glu Lys Met
          305             310             315             320
Asp Phe Ile Gly Glu Ala Ser Val Arg His Gly Phe Ser Thr Leu Asn
          325             330             335
Leu Asn Pro Gln Pro Gln Ala Ala Val Ala
          340             345

```

<210> SEQ ID NO 69

<211> LENGTH: 346

<212> TYPE: PRT

<213> ORGANISM: Prochlorococcus marinus

<400> SEQUENCE: 69

```

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys Arg
 1           5              10              15
Lys Ala Leu Gly Leu Gly Tyr Asp His Ile Ala Gln Gly Asp Leu Asp

```

-continued

20					25					30					
Val	Trp	Cys	Thr	Ala	Pro	Pro	Gln	Leu	Val	Glu	Asn	Val	Lys	Val	Val
		35					40					45			
Ser	Ala	Ile	Gly	Lys	Thr	Ile	Glu	Gly	Ala	Tyr	Ile	Asp	Ser	Cys	Phe
	50					55					60				
Val	Pro	Glu	Met	Leu	Ser	Arg	Phe	Lys	Thr	Ala	Arg	Arg	Lys	Val	Leu
65					70					75				80	
Asn	Ala	Met	Glu	Leu	Ala	Gln	Lys	Lys	Glu	Ile	Ser	Ile	Thr	Ala	Leu
				85					90					95	
Gly	Gly	Phe	Thr	Ser	Ile	Ile	Phe	Glu	Asn	Phe	Asn	Leu	Leu	Gln	Asn
			100						105					110	
Gln	Gln	Val	Arg	Asn	Thr	Thr	Leu	Asp	Trp	Gln	Arg	Phe	Thr	Thr	Gly
		115					120						125		
Asn	Thr	His	Thr	Ala	Trp	Val	Ile	Cys	Arg	Gln	Leu	Glu	Gln	Asn	Ala
	130					135					140				
Pro	Arg	Ile	Gly	Ile	Asp	Leu	Ser	Lys	Ser	Lys	Val	Ala	Val	Val	Gly
145					150					155					160
Ala	Thr	Gly	Asp	Ile	Gly	Ser	Ala	Val	Cys	Arg	Trp	Leu	Ser	Asn	Arg
			165						170					175	
Thr	Gly	Val	Ser	Glu	Leu	Leu	Leu	Val	Ala	Arg	Gln	Gln	Lys	Pro	Leu
			180					185						190	
Leu	Glu	Leu	Gln	Ser	Gln	Leu	Gly	Gly	Gly	Arg	Ile	Leu	Ser	Leu	Asp
		195					200					205			
Asp	Ala	Leu	Pro	Glu	Ala	Asp	Ile	Val	Ile	Trp	Val	Ala	Ser	Met	Pro
	210					215					220				
Lys	Thr	Leu	Glu	Ile	Asp	Pro	Ser	Lys	Ile	Lys	Arg	Pro	Cys	Leu	Met
225					230					235					240
Ile	Asp	Gly	Gly	Tyr	Pro	Lys	Asn	Leu	Gly	Glu	Lys	Phe	Ser	Gly	Pro
				245					250					255	
Gly	Ile	His	Val	Leu	Lys	Gly	Gly	Ile	Val	Gln	Phe	Phe	Lys	Asp	Ile
			260					265						270	
Gly	Trp	Ser	Met	Met	Glu	Leu	Ala	Glu	Met	Glu	Asn	Pro	Lys	Arg	Glu
		275					280					285			
Met	Phe	Ala	Cys	Phe	Ala	Glu	Ala	Met	Leu	Leu	Glu	Phe	Glu	Asn	Cys
	290					295					300				
His	Thr	Asn	Phe	Ser	Trp	Gly	Arg	Asn	Asn	Ile	Thr	Leu	Glu	Lys	Met
305					310					315					320
Asp	Phe	Ile	Gly	Lys	Ala	Ser	Glu	Arg	His	Gly	Phe	Ser	Ala	Val	Gly
				325					330					335	
Leu	Lys	Ser	Asn	Ile	Gln	Thr	Leu	Thr	Val						
			340					345							

<210> SEQ ID NO 70

<211> LENGTH: 346

<212> TYPE: PRT

<213> ORGANISM: Prochlorococcus marinus

<400> SEQUENCE: 70

Met	Phe	Gly	Leu	Ile	Gly	His	Ser	Thr	Ser	Phe	Glu	Asp	Ala	Lys	Arg
1				5					10					15	

Lys	Ala	Ser	Met	Leu	Gly	Phe	Asp	His	Ile	Ala	Asp	Gly	Asp	Leu	Asp
			20					25					30		

Val	Trp	Cys	Thr	Ala	Pro	Pro	Gln	Leu	Val	Glu	Asn	Val	Glu	Val	Lys
		35					40					45			

-continued

Ser Ala Thr Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys Phe
50 55 60

Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
65 70 75 80

Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala Leu
85 90 95

Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln His
100 105 110

Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr Gly
115 120 125

Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Ile Asn Ala
130 135 140

Pro Arg Ile Gly Ile Asp Leu Lys Lys Ala Thr Val Ala Val Ile Gly
145 150 155 160

Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn Lys
165 170 175

Thr Gly Ile Ser Glu Leu Leu Met Val Ala Arg Gln Gln Glu Pro Leu
180 185 190

Ala Leu Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Thr Thr Leu Asp
195 200 205

Lys Ala Leu Pro Gln Ala Asp Ile Val Val Trp Val Ala Ser Met Pro
210 215 220

Lys Thr Ile Glu Ile Asp Thr Asp Asn Leu Lys Lys Pro Cys Leu Met
225 230 235 240

Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly Glu
245 250 255

Asn Ile His Val Leu Lys Gly Gly Ile Val Glu Phe Phe Asn Asp Ile
260 265 270

Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg Glu
275 280 285

Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys Cys
290 295 300

His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys Met
305 310 315 320

Glu Phe Ile Gly Ala Ala Ser Leu Lys His Gly Phe Ser Ala Ile Gly
325 330 335

Leu Asp Lys Gln Pro Lys Val Leu Thr Val
340 345

<210> SEQ ID NO 71

<211> LENGTH: 348

<212> TYPE: PRT

<213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 71

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Ala Ala Arg Arg
1 5 10 15

Lys Ala Leu Glu Leu Gly Phe Asp His Ile Ala Glu Gly Asp Leu Asp
20 25 30

Val Trp Cys Ser Ala Pro Pro Gln Leu Val Glu His Leu Glu Val Thr
35 40 45

Ser Leu Thr Gly Lys Lys Ile Glu Gly Ala Tyr Ile Asp Ser Cys Phe
50 55 60

Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
65 70 75 80

-continued

Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala Leu
 85 90 95
 Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Lys His
 100 105 110
 Gln Thr Ile Arg Ser Thr Thr Leu Glu Trp Glu Arg Phe Thr Thr Gly
 115 120 125
 Asn Thr His Thr Ala Trp Val Ile Ser Arg Gln Val Glu Ile Asn Ala
 130 135 140
 Pro Leu Leu Gly Ile Asp Leu Ser Lys Ala Arg Val Ala Val Val Gly
 145 150 155 160
 Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Thr Gln Arg
 165 170 175
 Thr Gly Ile Lys Glu Leu Leu Met Val Ala Arg Gln Gln Gln Pro Leu
 180 185 190
 Lys Asp Leu Gln Gln Glu Leu Glu Gly Gly Arg Ile Leu Ser Leu Asp
 195 200 205
 Glu Ala Leu Pro Glu Ala Asp Val Val Val Trp Val Ala Ser Met Pro
 210 215 220
 Arg Thr Leu Glu Ile Asp Ser Asp Arg Leu Gln Lys Pro Cys Leu Met
 225 230 235 240
 Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Ser Arg Val Ala Gly Gln
 245 250 255
 Gly Val His Val Leu Lys Gly Gly Ile Val Glu Phe Val Ser Asp Ile
 260 265 270
 Gly Trp Thr Met Met Glu Asn Ala Glu Trp Gln Met Glu Lys Pro Gln
 275 280 285
 Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Ile Leu Leu Glu Phe Glu
 290 295 300
 Ala Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Thr Leu Glu
 305 310 315 320
 Lys Met Asp Phe Ile Gly Ala Ala Ser Val Arg His Gly Phe Ser Thr
 325 330 335
 Leu Asn Leu Gln Gly Gln Leu Gln Ala Ala Ala Ala
 340 345

<210> SEQ ID NO 72

<211> LENGTH: 346

<212> TYPE: PRT

<213> ORGANISM: Prochlorococcus marinus

<400> SEQUENCE: 72

Met Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys Arg
 1 5 10 15
 Lys Ala Ser Met Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu Asp
 20 25 30
 Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val Arg
 35 40 45
 Ser Ala Thr Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys Phe
 50 55 60
 Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val Leu
 65 70 75 80
 Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala Leu
 85 90 95
 Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln His

-continued

```

Asn Thr His Thr Ala Trp Val Ile Cys Lys Gln Leu Glu Ile Asn Ala
130                               135                               140

Pro Arg Ile Gly Ile Asp Leu Lys Lys Ala Thr Val Ala Val Ile Gly
145                               150                               155                               160

Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn Lys
165                               170                               175

Thr Gly Ile Ser Glu Leu Leu Met Val Ala Arg Gln Gln Glu Pro Leu
180                               185                               190

Ala Leu Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Thr Ser Leu Asp
195                               200                               205

Glu Ala Leu Pro Gln Ala Asp Ile Val Val Trp Val Ala Ser Met Pro
210                               215                               220

Lys Thr Ile Glu Ile Asn Thr Asp Asn Leu Gln Lys Pro Cys Leu Met
225                               230                               235                               240

Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly Glu
245                               250                               255

Asn Ile Tyr Val Leu Lys Gly Gly Ile Val Glu Phe Phe Asn Asp Ile
260                               265                               270

Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg Glu
275                               280                               285

Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys Cys
290                               295                               300

His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys Met
305                               310                               315                               320

Glu Phe Ile Gly Ala Ala Ser Leu Lys His Gly Phe Ser Ala Ile Gly
325                               330                               335

Leu Asp Lys Gln Pro Lys Val Leu Thr Val
340                               345

```

<210> SEQ ID NO 74

<211> LENGTH: 341

<212> TYPE: PRT

<213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 74

```

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Ala His Ala Lys Arg
1           5           10           15

Val Ala Asp Lys Leu Gly Tyr Ser Glu Tyr Ala Glu Ser Asp Leu Glu
20           25           30

Phe Trp Cys Met Ala Pro Pro Gln Val Val Asp Glu Ile Val Val Thr
35           40           45

Ser Ile Thr Gly Gln Lys Ile Tyr Gly Gln Tyr Val Glu Ser Cys Phe
50           55           60

Leu Pro Glu Met Leu Ala Gly Gly Arg Val Lys Ala Ala Cys Arg Lys
65           70           75           80

Ile Leu Asn Ala Met Ala Leu Ala Gln Arg Arg Gly Leu Asn Ile Thr
85           90           95

Thr Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Asn Phe Arg Leu Asp
100          105          110

Thr Leu Arg Arg Val Arg Asn Ile Asp Leu Glu Ile Arg Arg Phe Thr
115          120          125

Thr Gly Asn Thr His Thr Ala Tyr Ile Ile Cys Gln Gln Leu Gln Ala
130          135          140

Ala Ala Gln Arg Tyr Ala Met Asp Leu Ala Ala Ala Thr Val Ala Val
145          150          155          160

```

-continued

Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Ile Cys Gln Trp Leu Val
 165 170 175
 Ala His Thr Ser Pro Ala Lys Leu Leu Leu Ile Ala Arg Glu Arg Arg
 180 185 190
 Arg Leu Glu Glu Leu Gln Ala Lys Leu Lys Lys Gly Glu Val Cys Ser
 195 200 205
 Leu Glu Glu Ala Leu Pro Arg Ala Asp Phe Ile Val Trp Val Ala Ser
 210 215 220
 Met Ser Gln Gly Val Thr Leu Asp Pro Gln Val Leu Pro Asp Pro Cys
 225 230 235 240
 Val Ile Ile Asp Gly Gly Tyr Pro Lys Asn Ile Ala Ser Lys Leu Gln
 245 250 255
 Arg Lys Gly Leu Tyr Val Ile Asp Gly Gly Met Val Glu His Ser Leu
 260 265 270
 Asp Ile Glu Trp Asn Ile Met Gln Phe Leu Asn Val Ala Asn Pro Ala
 275 280 285
 Arg Gln Leu Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu
 290 295 300
 Gly Leu Tyr Thr Asn Phe Ser Trp Gly Arg Asn Leu Ile Thr Leu Glu
 305 310 315 320
 Lys Leu Asp Leu Ile Gly Gln Leu Ser Arg Lys His Gly Phe Arg Pro
 325 330 335
 Leu Met Pro Glu Ala
 340

<210> SEQ ID NO 75

<211> LENGTH: 341

<212> TYPE: PRT

<213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 75

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Ala His Ala Lys Arg
 1 5 10 15
 Val Ala Asp Lys Leu Gly Tyr Ser Glu Tyr Ala Glu Ser Asp Leu Glu
 20 25 30
 Phe Trp Cys Met Ala Pro Pro Gln Val Val Asp Glu Ile Thr Val Thr
 35 40 45
 Ser Ile Thr Gly Gln Lys Ile Tyr Gly Gln Tyr Val Glu Ser Cys Phe
 50 55 60
 Leu Pro Glu Met Leu Ala Gly Gly Arg Val Lys Ala Ala Cys Arg Lys
 65 70 75 80
 Val Leu Asn Ala Met Ala Leu Ala Gln Arg Arg Gly Leu Asn Ile Thr
 85 90 95
 Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Thr Phe Arg Leu Asp
 100 105 110
 Ser Leu Arg Arg Val Arg Asn Ile Asp Leu Glu Ile Gln Arg Phe Thr
 115 120 125
 Thr Gly Asn Thr His Thr Ala Tyr Ile Ile Cys Gln Gln Leu Gln Leu
 130 135 140
 Ala Ala Gln Arg Tyr Ala Met Asp Leu Ala Ala Ala Thr Val Ala Val
 145 150 155 160
 Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Ile Cys Gln Trp Leu Val
 165 170 175
 Ala His Thr His Leu Gly Lys Leu Leu Leu Ile Ala Arg Glu Arg Arg

-continued

Glu Ala Leu Pro Gln Ala Asp Ile Val Val Trp Val Ala Ser Met Pro
 210 215 220
 Lys Thr Ile Glu Ile Glu Ile Glu Asn Leu Lys Lys Pro Cys Leu Met
 225 230 235 240
 Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Lys Gly Lys
 245 250 255
 Asn Ile His Val Leu Lys Gly Gly Ile Val Glu Phe Phe Asn Asp Ile
 260 265 270
 Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg Glu
 275 280 285
 Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys Cys
 290 295 300
 His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys Met
 305 310 315 320
 Glu Phe Ile Gly Ala Ala Ser Leu Lys His Gly Phe Ser Ala Ile Gly
 325 330 335
 Leu Asp Lys Gln Pro Lys Val Leu Thr Val
 340 345

<210> SEQ ID NO 77

<211> LENGTH: 231

<212> TYPE: PRT

<213> ORGANISM: *Synechococcus elongatus*

<400> SEQUENCE: 77

Met Pro Gln Leu Glu Ala Ser Leu Glu Leu Asp Phe Gln Ser Glu Ser
 1 5 10 15
 Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu
 20 25 30
 Gln Glu Ala Phe Asp Asn Tyr Asn Arg Leu Ala Glu Met Leu Pro Asp
 35 40 45
 Gln Arg Asp Glu Leu His Lys Leu Ala Lys Met Glu Gln Arg His Met
 50 55 60
 Lys Gly Phe Met Ala Cys Gly Lys Asn Leu Ser Val Thr Pro Asp Met
 65 70 75 80
 Gly Phe Ala Gln Lys Phe Phe Glu Arg Leu His Glu Asn Phe Lys Ala
 85 90 95
 Ala Ala Ala Glu Gly Lys Val Val Thr Cys Leu Leu Ile Gln Ser Leu
 100 105 110
 Ile Ile Glu Cys Phe Ala Ile Ala Ala Tyr Asn Ile Tyr Ile Pro Val
 115 120 125
 Ala Asp Ala Phe Ala Arg Lys Ile Thr Glu Gly Val Val Arg Asp Glu
 130 135 140
 Tyr Leu His Arg Asn Phe Gly Glu Glu Trp Leu Lys Ala Asn Phe Asp
 145 150 155 160
 Ala Ser Lys Ala Glu Leu Glu Glu Ala Asn Arg Gln Asn Leu Pro Leu
 165 170 175
 Val Trp Leu Met Leu Asn Glu Val Ala Asp Asp Ala Arg Glu Leu Gly
 180 185 190
 Met Glu Arg Glu Ser Leu Val Glu Asp Phe Met Ile Ala Tyr Gly Glu
 195 200 205
 Ala Leu Glu Asn Ile Gly Phe Thr Thr Arg Glu Ile Met Arg Met Ser
 210 215 220
 Ala Tyr Gly Leu Ala Ala Val
 225 230

-continued

```

<210> SEQ ID NO 78
<211> LENGTH: 254
<212> TYPE: PRT
<213> ORGANISM: Synechococcus elongatus

<400> SEQUENCE: 78

Met Arg Thr Pro Trp Asp Pro Pro Asn Pro Thr Phe Ser Leu Ser Ser
1          5          10          15
Val Ser Gly Asp Arg Arg Leu Met Pro Gln Leu Glu Ala Ser Leu Glu
20          25          30
Leu Asp Phe Gln Ser Glu Ser Tyr Lys Asp Ala Tyr Ser Arg Ile Asn
35          40          45
Ala Ile Val Ile Glu Gly Glu Gln Glu Ala Phe Asp Asn Tyr Asn Arg
50          55          60
Leu Ala Glu Met Leu Pro Asp Gln Arg Asp Glu Leu His Lys Leu Ala
65          70          75          80
Lys Met Glu Gln Arg His Met Lys Gly Phe Met Ala Cys Gly Lys Asn
85          90          95
Leu Ser Val Thr Pro Asp Met Gly Phe Ala Gln Lys Phe Phe Glu Arg
100         105         110
Leu His Glu Asn Phe Lys Ala Ala Ala Ala Glu Gly Lys Val Val Thr
115         120         125
Cys Leu Leu Ile Gln Ser Leu Ile Ile Glu Cys Phe Ala Ile Ala Ala
130         135         140
Tyr Asn Ile Tyr Ile Pro Val Ala Asp Ala Phe Ala Arg Lys Ile Thr
145         150         155         160
Glu Gly Val Val Arg Asp Glu Tyr Leu His Arg Asn Phe Gly Glu Glu
165         170         175
Trp Leu Lys Ala Asn Phe Asp Ala Ser Lys Ala Glu Leu Glu Glu Ala
180         185         190
Asn Arg Gln Asn Leu Pro Leu Val Trp Leu Met Leu Asn Glu Val Ala
195         200         205
Asp Asp Ala Arg Glu Leu Gly Met Glu Arg Glu Ser Leu Val Glu Asp
210         215         220
Phe Met Ile Ala Tyr Gly Glu Ala Leu Glu Asn Ile Gly Phe Thr Thr
225         230         235         240
Arg Glu Ile Met Arg Met Ser Ala Tyr Gly Leu Ala Ala Val
245         250

```

```

<210> SEQ ID NO 79
<211> LENGTH: 231
<212> TYPE: PRT
<213> ORGANISM: Arthrospira maxima

<400> SEQUENCE: 79

Met Pro Gln Leu Glu Thr Ile Thr Glu Leu Asp Phe Gln Asn Glu Thr
1          5          10          15
Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu
20          25          30
Gln Glu Ala Ser Asp Asn Tyr Ile Lys Leu Gly Glu Met Leu Pro Glu
35          40          45
Glu Arg Glu Glu Leu Ile Arg Leu Ser Lys Met Glu Lys Arg His Lys
50          55          60
Lys Gly Phe Gln Ala Cys Gly Arg Asn Leu Glu Val Thr Pro Asp Met
65          70          75          80

```


-continued

210	215	220
Thr Arg Asp Ile Met	Arg Met Ser Ala Tyr Gly	Leu Thr Ala Ala
225	230	235
<210> SEQ ID NO 81		
<211> LENGTH: 231		
<212> TYPE: PRT		
<213> ORGANISM: Lyngbya sp.		
<400> SEQUENCE: 81		
Met Pro Gln Leu Glu Ala Ile Ala Glu Ile Asp Phe Asn Thr Asn Thr		
1	5	10 15
Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu		
	20	25 30
Gln Val Ala His Asp Asn Tyr Ile Lys Leu Gly Glu Met Leu Pro Asp		
	35	40 45
Gln Lys Asp Glu Leu Val Arg Leu Ser Lys Met Glu Lys Arg His Met		
	50	55 60
Lys Gly Phe Gln Ala Cys Gly Arg Asn Leu Glu Val Thr Ala Asp Met		
	65	70 75 80
Asp Tyr Ala His Gln Phe Phe Ser Gln Leu His Gln Asn Phe Lys Asp		
	85	90 95
Ala Ala Ala Gln Gly Lys Val Val Thr Cys Leu Leu Ile Gln Ser Leu		
	100	105 110
Ile Ile Glu Ser Phe Ala Ile Ala Ala Tyr Asn Ile Tyr Ile Pro Val		
	115	120 125
Ala Asp Pro Phe Ala Arg Lys Ile Thr Glu Gly Val Val Asp Asp Glu		
	130	135 140
Tyr Met His Leu Asn Phe Gly Glu Glu Trp Leu Lys Ala His Phe Glu		
	145	150 155 160
Glu Ser Lys Ala Glu Leu Gln Glu Ala Asn Ser Gln Asn Leu Pro Leu		
	165	170 175
Val Trp Lys Met Leu Asn Glu Val Glu Asn Asp Ala His Ile Leu Gly		
	180	185 190
Met Glu Lys Asp Ala Leu Val Glu Asp Phe Met Ile Ala Tyr Gly Glu		
	195	200 205
Ala Leu Asn Asn Ile Gly Phe Thr Thr Arg Glu Ile Met Arg Met Ser		
	210	215 220
Ala His Gly Leu Thr Thr Ala		
225	230	

<210> SEQ ID NO 82
 <211> LENGTH: 231
 <212> TYPE: PRT
 <213> ORGANISM: Nodularia spumigena

<400> SEQUENCE: 82

Met Gln Gln Leu Ala Ala Glu Leu Lys Ile Asp Phe Gln Ser Glu Lys		
1	5	10 15
Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu		
	20	25 30
Gln Glu Ala His Asp Asn Tyr Ile Thr Leu Gly Glu Met Leu Pro Glu		
	35	40 45
Leu Lys Asp Glu Leu Ile Arg Leu Ser Lys Met Glu Ser Arg His Lys		
	50	55 60
Lys Gly Phe Glu Ala Cys Gly Arg Asn Leu Ser Val Lys Pro Asp Met		

-continued

```

65              70              75              80
Pro Phe Ala Gln Lys Phe Phe Ser Gly Leu His Glu Asn Phe Gln Lys
      85              90              95
Ala Ala Ala Glu Gly Gln Val Val Thr Cys Leu Leu Ile Gln Ser Leu
      100             105             110
Ile Ile Glu Cys Phe Ala Ile Ala Ala Tyr Asn Ile Tyr Ile Pro Val
      115             120             125
Ala Asp Asp Phe Ala Arg Lys Ile Thr Glu Gly Val Val Lys Asp Glu
      130             135             140
Tyr Ser His Leu Asn Phe Gly Glu Val Trp Leu Lys Glu Asn Phe Ala
      145             150             155             160
Gln Ser Lys Ala Glu Leu Glu Ala Ala Asn Arg Gln Asn Leu Pro Ile
      165             170             175
Val Trp Lys Met Leu Asn Glu Val Glu Asn Asp Ala His Val Leu Ala
      180             185             190
Met Glu Lys Glu Ala Leu Val Glu Asp Phe Met Ile Gln Tyr Gly Glu
      195             200             205
Thr Leu Ser Asn Ile Gly Phe Thr Thr Arg Asp Ile Met Lys Met Ser
      210             215             220
Ala Tyr Gly Leu Thr Ala Ala
      225             230

```

<210> SEQ ID NO 83

<211> LENGTH: 231

<212> TYPE: PRT

<213> ORGANISM: Microcystis aeruginosa

<400> SEQUENCE: 83

```

Met Pro Glu Leu Ala Val Pro Leu Glu Leu Asp Phe Thr Ser Glu Thr
  1              5              10              15
Tyr Lys Ser Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu
      20              25              30
Tyr Glu Ala Asn Ser Asn Tyr Ile Gln Leu Ala Asp Ile Leu Thr Asp
      35              40              45
Asn Lys Glu Glu Leu His Arg Leu Ala Lys Met Glu Asn Arg His Met
      50              55              60
Lys Gly Phe Gln Ala Cys Gly Gln Asn Leu Lys Ile Thr Pro Asp Met
      65              70              75              80
Asp Tyr Ala Arg Glu Phe Phe Ser Ser Leu His Asn Asn Phe Gln Ile
      85              90              95
Ala Tyr Ala Glu Gly Lys Val Val Thr Cys Leu Leu Ile Gln Ser Leu
      100             105             110
Ile Ile Glu Ala Phe Ala Ile Ala Ala Tyr Asn Ile Tyr Ile Pro Val
      115             120             125
Ala Asp Pro Phe Ala Arg Lys Ile Thr Glu Gly Val Val Lys Asp Glu
      130             135             140
Tyr Leu His Leu Asn Phe Gly Glu Glu Trp Leu Lys Ala Asn Phe Glu
      145             150             155             160
Thr Ala Lys Glu Glu Leu Glu Ala Ala Asn Arg Ala Asn Leu Pro Ile
      165             170             175
Val Trp Arg Met Leu Asn Gln Val Glu Asn Asp Ala Arg Val Leu Gly
      180             185             190
Met Glu Lys Glu Ala Leu Val Glu Asp Phe Met Ile Ser Tyr Gly Glu
      195             200             205

```

-continued

Ala Leu Ser Asn Ile Gly Phe Ser Thr Arg Asp Ile Met Arg Met Ser
210 215 220

Ala Tyr Gly Leu Thr Ala Val
225 230

<210> SEQ ID NO 84
<211> LENGTH: 231
<212> TYPE: PRT
<213> ORGANISM: *Microcystis aeruginosa*

<400> SEQUENCE: 84

Met Pro Glu Leu Ala Val Pro Leu Glu Leu Asp Phe Thr Ser Glu Thr
1 5 10 15

Tyr Lys Ser Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu
20 25 30

Tyr Glu Ala Asn Ser Asn Tyr Ile Gln Leu Ala Asp Ile Leu Thr Asp
35 40 45

Asn Lys Glu Glu Leu His Arg Leu Ala Lys Met Glu Asn Arg His Met
50 55 60

Lys Gly Phe Gln Ala Cys Gly Gln Asn Leu Gln Ile Thr Pro Asp Met
65 70 75 80

Glu Tyr Ala Lys Glu Phe Phe Ser Ser Leu His Asn Asn Phe Gln Ile
85 90 95

Ala Tyr Ala Glu Gly Lys Val Val Thr Cys Leu Leu Ile Gln Ser Leu
100 105 110

Ile Ile Glu Ala Phe Ala Ile Ala Ala Tyr Asn Ile Tyr Ile Pro Val
115 120 125

Ala Asp Pro Phe Ala Arg Lys Ile Thr Glu Ser Val Val Lys Asp Glu
130 135 140

Tyr Leu His Leu Asn Phe Gly Glu Glu Trp Leu Lys Ala Asn Phe Glu
145 150 155 160

Thr Ala Lys Glu Glu Leu Glu Ala Ala Asn Arg Ala Asn Leu Pro Ile
165 170 175

Val Trp Arg Met Leu Asn Gln Val Glu Asp Asp Ala Arg Val Leu Ala
180 185 190

Met Glu Lys Glu Ala Leu Val Glu Asp Phe Met Ile Ser Tyr Gly Glu
195 200 205

Ala Leu Asn Asn Ile Gly Phe Ser Thr Arg Asp Ile Met Arg Met Ser
210 215 220

Ala Tyr Gly Leu Thr Ala Val
225 230

<210> SEQ ID NO 85
<211> LENGTH: 231
<212> TYPE: PRT
<213> ORGANISM: *Nostoc sp.*

<400> SEQUENCE: 85

Met Gln Gln Val Ala Ala Asp Leu Glu Ile Asp Phe Lys Ser Glu Lys
1 5 10 15

Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu
20 25 30

Gln Glu Ala Tyr Glu Asn Tyr Ile Gln Leu Ser Gln Leu Leu Pro Asp
35 40 45

Asp Lys Glu Asp Leu Ile Arg Leu Ser Lys Met Glu Ser Arg His Lys
50 55 60

-continued

Lys Gly Phe Glu Ala Cys Gly Arg Asn Leu Gln Val Ser Pro Asp Met
 65 70 75 80
 Glu Phe Ala Lys Glu Phe Phe Ala Gly Leu His Gly Asn Phe Gln Lys
 85 90 95
 Ala Ala Ala Glu Gly Lys Ile Val Thr Cys Leu Leu Ile Gln Ser Leu
 100 105 110
 Ile Ile Glu Cys Phe Ala Ile Ala Ala Tyr Asn Ile Tyr Ile Pro Val
 115 120 125
 Ala Asp Asp Phe Ala Arg Lys Ile Thr Glu Gly Val Val Lys Asp Glu
 130 135 140
 Tyr Ser His Leu Asn Phe Gly Glu Val Trp Leu Gln Lys Asn Phe Ala
 145 150 155 160
 Gln Ser Lys Ala Glu Leu Glu Glu Ala Asn Arg His Asn Leu Pro Ile
 165 170 175
 Val Trp Lys Met Leu Asn Gln Val Ala Asp Asp Ala Ala Val Leu Ala
 180 185 190
 Met Glu Lys Glu Ala Leu Val Glu Asp Phe Met Ile Gln Tyr Gly Glu
 195 200 205
 Ala Leu Ser Asn Ile Gly Phe Thr Thr Arg Asp Ile Met Arg Met Ser
 210 215 220
 Ala Tyr Gly Leu Thr Ala Ala
 225 230

<210> SEQ ID NO 86

<211> LENGTH: 231

<212> TYPE: PRT

<213> ORGANISM: Anabaena variabilis

<400> SEQUENCE: 86

Met Gln Gln Val Ala Ala Asp Leu Glu Ile Asp Phe Lys Ser Glu Lys
 1 5 10 15
 Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu
 20 25 30
 Gln Glu Ala Tyr Glu Asn Tyr Ile Gln Leu Ser Gln Leu Leu Pro Asp
 35 40 45
 Asp Lys Glu Asp Leu Ile Arg Leu Ser Lys Met Glu Ser Arg His Lys
 50 55 60
 Lys Gly Phe Glu Ala Cys Gly Arg Asn Leu Gln Val Ser Pro Asp Ile
 65 70 75 80
 Glu Phe Ala Lys Glu Phe Phe Ala Gly Leu His Gly Asn Phe Gln Lys
 85 90 95
 Ala Ala Ala Glu Gly Lys Val Val Thr Cys Leu Leu Ile Gln Ser Leu
 100 105 110
 Ile Ile Glu Cys Phe Ala Ile Ala Ala Tyr Asn Ile Tyr Ile Pro Val
 115 120 125
 Ala Asp Asp Phe Ala Arg Lys Ile Thr Glu Gly Val Val Lys Asp Glu
 130 135 140
 Tyr Ser His Leu Asn Phe Gly Glu Val Trp Leu Gln Lys Asn Phe Ala
 145 150 155 160
 Gln Ser Lys Ala Glu Leu Glu Glu Ala Asn Arg His Asn Leu Pro Ile
 165 170 175
 Val Trp Lys Met Leu Asn Gln Val Ala Asp Asp Ala Ala Val Leu Ala
 180 185 190
 Met Glu Lys Glu Ala Leu Val Glu Asp Phe Met Ile Gln Tyr Gly Glu
 195 200 205

-continued

Ala Leu Ser Asn Ile Gly Phe Thr Thr Arg Asp Ile Met Arg Met Ser
 210 215 220

Ala Tyr Gly Leu Thr Ala Ala
 225 230

<210> SEQ ID NO 87
 <211> LENGTH: 231
 <212> TYPE: PRT
 <213> ORGANISM: *Crocospaera watsonii*

<400> SEQUENCE: 87

Met Gln Glu Leu Ala Val Arg Ser Glu Leu Asp Phe Asn Ser Glu Thr
 1 5 10 15

Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu
 20 25 30

Gln Glu Ala Tyr Glu Asn Tyr Ile Asp Met Gly Glu Leu Leu Pro Gly
 35 40 45

Asp Lys Asp Glu Leu Ile Arg Leu Ser Lys Met Glu Asn Arg His Lys
 50 55 60

Lys Gly Phe Gln Ala Cys Gly Lys Asn Leu Lys Val Thr Pro Asp Met
 65 70 75 80

Asp Tyr Ala Glu Arg Phe Phe Ser Gln Leu His Gly Asn Phe Gln Thr
 85 90 95

Ala Lys Ala Glu Gly Lys Ile Val Thr Cys Leu Leu Ile Gln Ser Leu
 100 105 110

Ile Ile Glu Ala Phe Ala Ile Ala Ala Tyr Asn Ile Tyr Ile Pro Val
 115 120 125

Ala Asp Pro Phe Ala Arg Lys Ile Thr Glu Asn Val Val Lys Asp Glu
 130 135 140

Tyr Ser His Leu Asn Phe Gly Glu Val Trp Leu Lys Glu Asn Phe Glu
 145 150 155 160

Ala Ser Lys Ala Glu Leu Glu Gln Ala Asn Lys Glu Asn Leu Pro Ile
 165 170 175

Val Trp Gln Met Leu Asn Glu Val Glu Asp Asp Ala Glu Ile Leu Gly
 180 185 190

Met Glu Lys Glu Ala Leu Val Glu Asp Phe Met Ile Ser Tyr Gly Glu
 195 200 205

Ala Leu Gly Asn Ile Gly Phe Ser Thr Arg Glu Ile Met Lys Met Ser
 210 215 220

Ala His Gly Leu Ala Ala Val
 225 230

<210> SEQ ID NO 88
 <211> LENGTH: 251
 <212> TYPE: PRT
 <213> ORGANISM: *Trichodesmium erythraeum*

<400> SEQUENCE: 88

Met Pro Lys Leu Glu Ile Ile Pro Thr Met Asp Ser Gln Ser Glu Thr
 1 5 10 15

Lys Leu Glu Lys Val Lys Ser Gln Ser Glu Gly Asp Gln Ile Asn Phe
 20 25 30

Glu Thr Glu Thr Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val
 35 40 45

Ile Glu Gly Glu Gln Glu Ala Tyr Lys Asn Tyr Ile Lys Leu Ala Glu
 50 55 60

-continued

Met Leu Pro Asp Glu Lys Asp Glu Leu Ile Lys Leu Ser Lys Met Glu
65 70 75 80

Asn Arg His Lys Lys Gly Phe Glu Ala Cys Gly Arg Asn Leu His Val
85 90 95

Thr Pro Asp Met Glu Phe Ala Lys Lys Phe Phe Glu Pro Leu His Glu
100 105 110

Asn Phe Gln Thr Ala Ala Ala Thr Gly Asn Val Val Thr Cys Leu Leu
115 120 125

Ile Gln Ser Leu Ile Ile Glu Cys Phe Ala Ile Ala Ala Tyr Asn Ile
130 135 140

Tyr Ile Pro Val Ala Asp Pro Phe Ala Arg Lys Ile Thr Glu Ser Val
145 150 155 160

Val Lys Asp Glu Tyr Ser His Leu Asn Phe Gly Glu Val Trp Leu Lys
165 170 175

Glu Tyr Phe Glu Asp Ser Lys Gln Glu Leu Gln Lys Ala Asn Arg Gln
180 185 190

Asn Leu Pro Leu Val Trp Lys Met Leu Asn Gln Val Glu Lys Asp Ala
195 200 205

Lys Thr Leu Glu Met Glu Lys Glu Ala Leu Ile Glu Asp Phe Met Ile
210 215 220

Ala Tyr Gly Glu Ala Leu Asn Asn Ile Gly Phe Thr Thr Gly Glu Ile
225 230 235 240

Met Arg Met Ser Ala Tyr Gly Leu Ile Ala Ala
245 250

<210> SEQ ID NO 89

<211> LENGTH: 231

<212> TYPE: PRT

<213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 89

Met Gln Thr Leu Glu Val Ser Pro Ala Met Asp Phe Gln Ser Glu Thr
1 5 10 15

Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu
20 25 30

Leu Glu Ala Asn Asn Asn Tyr Lys Gln Leu Ser Glu His Leu Gly Asp
35 40 45

Phe Lys Asp Asp Leu Leu Lys Leu Ala Arg Met Glu Asn Arg His Met
50 55 60

Lys Gly Phe Gln Ala Cys Gly Lys Asn Leu Ser Val Asn Pro Asp Met
65 70 75 80

Pro Phe Ala Lys Glu Phe Phe Ala Gln Leu His Asp Asn Phe Gln Thr
85 90 95

Ala Leu Ala Glu Gly Lys Ile Val Thr Cys Leu Leu Ile Gln Ser Leu
100 105 110

Ile Ile Glu Thr Phe Ala Ile Ser Ala Tyr Asn Ile Tyr Ile Pro Val
115 120 125

Ala Asp Asp Phe Ala Arg Lys Ile Thr Glu Gly Val Val Lys Asp Glu
130 135 140

Tyr Met His Leu Asn Phe Gly Glu Glu Trp Leu Lys Ala Asn Phe Glu
145 150 155 160

Ala Ser Lys Ala Glu Leu Glu Thr Ala Asn Arg Ala Asn Leu Pro Leu
165 170 175

Ile Trp Lys Met Leu Asn Gln Val Glu Glu Asp Ala Ala Val Leu Gly

-continued

```

180          185          190
Met Glu Lys Asp Ala Leu Ile Glu Asp Phe Met Ile Thr Tyr Gly Glu
   195                200                205

Ala Leu Ala Asn Ile Gly Phe Ser Ala Arg Asp Val Met Arg Leu Ser
   210                215                220

Ala Gln Gly Leu Ala Ala Val
   225                230

```

```

<210> SEQ ID NO 90
<211> LENGTH: 232
<212> TYPE: PRT
<213> ORGANISM: Nostoc azollae

```

```

<400> SEQUENCE: 90

```

```

Met Gln Gln Leu Val Glu Glu Ile Glu Lys Ile Asp Phe Gln Ser Glu
 1          5          10          15

Lys Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly
   20          25          30

Glu Gln Glu Ala His Glu Asn Tyr Ile Thr Leu Ala Lys Leu Leu Pro
   35          40          45

Glu Ser Lys Glu Glu Leu Met Arg Leu Ser Lys Met Glu Ser Arg His
   50          55          60

Lys Lys Gly Phe Glu Ala Cys Gly Arg Asn Leu Gln Val Thr Pro Asp
 65          70          75          80

Met Gln Phe Ala Lys Glu Phe Phe Ser Gly Leu His Gln Asn Phe Gln
   85          90          95

Thr Ala Ala Ala Ala Gly Asn Val Val Thr Cys Leu Leu Ile Gln Ser
   100         105         110

Leu Ile Ile Glu Cys Phe Ala Ile Ala Ala Tyr Asn Ile Tyr Ile Pro
   115         120         125

Val Ala Asp Asp Phe Ala Arg Lys Ile Thr Glu Gly Val Val Lys Glu
   130         135         140

Glu Tyr Ser His Leu Asn Phe Gly Glu Val Trp Leu Lys Glu His Phe
 145         150         155         160

Ala Glu Ser Lys Ala Glu Leu Asp Asp Ala Asn Arg Gln Asn Leu Pro
   165         170         175

Ile Val Trp Gln Met Leu Asn Gln Val Ala Asp Asp Ala Arg Val Leu
   180         185         190

Ala Met Glu Lys Glu Ala Leu Val Glu Asp Phe Met Ile Gln Tyr Gly
   195         200         205

Glu Ala Leu Ser Asn Ile Gly Phe Thr Thr Arg Asp Ile Ile Arg Leu
   210         215         220

Ser Ala Tyr Gly Leu Ala Thr Val
 225         230

```

```

<210> SEQ ID NO 91
<211> LENGTH: 231
<212> TYPE: PRT
<213> ORGANISM: Synechocystis sp.

```

```

<400> SEQUENCE: 91

```

```

Met Pro Glu Leu Ala Val Arg Thr Glu Phe Asp Tyr Ser Ser Glu Ile
 1          5          10          15

Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu
   20          25          30

Gln Glu Ala Tyr Ser Asn Tyr Leu Gln Met Ala Glu Leu Leu Pro Glu

```

-continued

35					40					45					
Asp	Lys	Glu	Glu	Leu	Thr	Arg	Leu	Ala	Lys	Met	Glu	Asn	Arg	His	Lys
50						55					60				
Lys	Gly	Phe	Gln	Ala	Cys	Gly	Asn	Asn	Leu	Gln	Val	Asn	Pro	Asp	Met
65					70					75					80
Pro	Tyr	Ala	Gln	Glu	Phe	Phe	Ala	Gly	Leu	His	Gly	Asn	Phe	Gln	His
				85					90					95	
Ala	Phe	Ser	Glu	Gly	Lys	Val	Val	Thr	Cys	Leu	Leu	Ile	Gln	Ala	Leu
			100					105					110		
Ile	Ile	Glu	Ala	Phe	Ala	Ile	Ala	Ala	Tyr	Asn	Ile	Tyr	Ile	Pro	Val
		115					120					125			
Ala	Asp	Asp	Phe	Ala	Arg	Lys	Ile	Thr	Glu	Gly	Val	Val	Lys	Asp	Glu
130					135						140				
Tyr	Thr	His	Leu	Asn	Tyr	Gly	Glu	Glu	Trp	Leu	Lys	Ala	Asn	Phe	Ala
145				150						155					160
Thr	Ala	Lys	Glu	Glu	Leu	Glu	Gln	Ala	Asn	Lys	Glu	Asn	Leu	Pro	Leu
			165						170					175	
Val	Trp	Lys	Met	Leu	Asn	Gln	Val	Gln	Gly	Asp	Ala	Lys	Val	Leu	Gly
			180					185						190	
Met	Glu	Lys	Glu	Ala	Leu	Val	Glu	Asp	Phe	Met	Ile	Ser	Tyr	Gly	Glu
		195					200					205			
Ala	Leu	Ser	Asn	Ile	Gly	Phe	Ser	Thr	Arg	Glu	Ile	Met	Arg	Met	Ser
210					215						220				
Ser	Tyr	Gly	Leu	Ala	Gly	Val									
225					230										

<210> SEQ ID NO 92

<211> LENGTH: 231

<212> TYPE: PRT

<213> ORGANISM: Cyanothecce sp.

<400> SEQUENCE: 92

Met	Gln	Glu	Leu	Ala	Leu	Arg	Ser	Glu	Leu	Asp	Phe	Asn	Ser	Glu	Thr
1				5					10					15	
Tyr	Lys	Asp	Ala	Tyr	Ser	Arg	Ile	Asn	Ala	Ile	Val	Ile	Glu	Gly	Glu
			20					25					30		
Gln	Glu	Ala	Tyr	Gln	Asn	Tyr	Leu	Asp	Met	Ala	Gln	Leu	Leu	Pro	Glu
		35					40					45			
Asp	Glu	Ala	Glu	Leu	Ile	Arg	Leu	Ser	Lys	Met	Glu	Asn	Arg	His	Lys
50					55						60				
Lys	Gly	Phe	Gln	Ala	Cys	Gly	Lys	Asn	Leu	Asn	Val	Thr	Pro	Asp	Met
65					70					75					80
Asp	Tyr	Ala	Gln	Gln	Phe	Phe	Ala	Glu	Leu	His	Gly	Asn	Phe	Gln	Lys
				85					90					95	
Ala	Lys	Ala	Glu	Gly	Lys	Ile	Val	Thr	Cys	Leu	Leu	Ile	Gln	Ser	Leu
			100					105					110		
Ile	Ile	Glu	Ala	Phe	Ala	Ile	Ala	Ala	Tyr	Asn	Ile	Tyr	Ile	Pro	Val
		115					120					125			
Ala	Asp	Pro	Phe	Ala	Arg	Lys	Ile	Thr	Glu	Gly	Val	Val	Lys	Asp	Glu
130					135						140				
Tyr	Thr	His	Leu	Asn	Phe	Gly	Glu	Val	Trp	Leu	Lys	Glu	His	Phe	Glu
145				150						155					160
Ala	Ser	Lys	Ala	Glu	Leu	Glu	Asp	Ala	Asn	Lys	Glu	Asn	Leu	Pro	Leu
				165					170					175	

-continued

Val Trp Gln Met Leu Asn Gln Val Glu Lys Asp Ala Glu Val Leu Gly
 180 185 190

Met Glu Lys Glu Ala Leu Val Glu Asp Phe Met Ile Ser Tyr Gly Glu
 195 200 205

Ala Leu Ser Asn Ile Gly Phe Ser Thr Arg Glu Ile Met Lys Met Ser
 210 215 220

Ala Tyr Gly Leu Arg Ala Ala
 225 230

<210> SEQ ID NO 93
 <211> LENGTH: 231
 <212> TYPE: PRT
 <213> ORGANISM: Cyanothecce sp.

<400> SEQUENCE: 93

Met Gln Glu Leu Ala Leu Arg Ser Glu Leu Asp Phe Asn Ser Glu Thr
 1 5 10 15

Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu
 20 25 30

Gln Glu Ala His Gln Asn Tyr Ile Asp Met Ala Gln Leu Leu Pro Glu
 35 40 45

Asp Glu Ala Glu Leu Ile Arg Leu Ser Lys Met Glu Asn Arg His Lys
 50 55 60

Lys Gly Phe Gln Ala Cys Gly Lys Asn Leu Asp Val Thr Pro Asp Met
 65 70 75 80

Asp Tyr Ala Gln Gln Phe Phe Ser Gln Leu His Asn Asn Phe Gln Thr
 85 90 95

Ala Lys Ala Glu Gly Lys Ile Val Thr Cys Leu Leu Ile Gln Ser Leu
 100 105 110

Ile Ile Glu Ala Phe Ala Ile Ala Ala Tyr Asn Ile Tyr Ile Pro Val
 115 120 125

Ala Asp Pro Phe Ala Arg Lys Ile Thr Glu Gly Val Val Lys Asp Glu
 130 135 140

Tyr Thr His Leu Asn Phe Gly Glu Ile Trp Leu Lys Glu His Phe Glu
 145 150 155 160

Ala Ser Lys Ala Glu Leu Glu Glu Ala Asn Lys Lys Asn Leu Pro Ile
 165 170 175

Val Trp Gln Met Leu Asn Gln Val Glu Lys Asp Ala Glu Val Leu Gly
 180 185 190

Met Glu Lys Glu Ala Leu Val Glu Asp Phe Met Ile Ser Tyr Gly Glu
 195 200 205

Ala Leu Ser Asn Ile Gly Phe Ser Thr Arg Glu Ile Met Lys Met Ser
 210 215 220

Ser His Gly Leu Ser Ala Ala
 225 230

<210> SEQ ID NO 94
 <211> LENGTH: 231
 <212> TYPE: PRT
 <213> ORGANISM: Cyanothecce sp.

<400> SEQUENCE: 94

Met Pro Gln Val Gln Ser Pro Ser Ala Ile Asp Phe Tyr Ser Glu Thr
 1 5 10 15

Tyr Gln Asp Ala Tyr Ser Arg Ile Asp Ala Ile Val Ile Glu Gly Glu
 20 25 30

-continued

Ile Val Trp Lys Met Leu Asn Gln Val Glu Gly Asp Ala His Thr Met
 180 185 190
 Ala Met Glu Lys Asp Ala Leu Val Glu Asp Phe Met Ile Gln Tyr Gly
 195 200 205
 Glu Ala Leu Ser Asn Ile Gly Phe Ser Thr Arg Asp Ile Met Arg Leu
 210 215 220
 Ser Ala Tyr Gly Leu Ile Gly Ala
 225 230

<210> SEQ ID NO 96
 <211> LENGTH: 231
 <212> TYPE: PRT
 <213> ORGANISM: *Acaryochloris marina*

<400> SEQUENCE: 96

Met Pro Gln Thr Gln Ala Ile Ser Glu Ile Asp Phe Tyr Ser Asp Thr
 1 5 10 15
 Tyr Lys Asp Ala Tyr Ser Arg Ile Asp Gly Ile Val Ile Glu Gly Glu
 20 25 30
 Gln Glu Ala His Glu Asn Tyr Ile Arg Leu Gly Glu Met Leu Pro Glu
 35 40 45
 His Gln Asp Asp Phe Ile Arg Leu Ser Lys Met Glu Ala Arg His Lys
 50 55 60
 Lys Gly Phe Glu Ala Cys Gly Arg Asn Leu Lys Val Thr Cys Asp Leu
 65 70 75 80
 Asp Phe Ala Arg Arg Phe Phe Ser Asp Leu His Lys Asn Phe Gln Asp
 85 90 95
 Ala Ala Ala Glu Asp Lys Val Pro Thr Cys Leu Val Ile Gln Ser Leu
 100 105 110
 Ile Ile Glu Cys Phe Ala Ile Ala Ala Tyr Asn Ile Tyr Ile Pro Val
 115 120 125
 Ala Asp Asp Phe Ala Arg Lys Ile Thr Glu Ser Val Val Lys Asp Glu
 130 135 140
 Tyr Gln His Leu Asn Tyr Gly Glu Glu Trp Leu Lys Ala His Phe Asp
 145 150 155 160
 Asp Val Lys Ala Glu Ile Gln Glu Ala Asn Arg Lys Asn Leu Pro Ile
 165 170 175
 Val Trp Arg Met Leu Asn Glu Val Asp Lys Asp Ala Ala Val Leu Gly
 180 185 190
 Met Glu Lys Glu Ala Leu Val Glu Asp Phe Met Ile Gln Tyr Gly Glu
 195 200 205
 Ala Leu Ser Asn Ile Gly Phe Ser Thr Gly Glu Ile Met Arg Met Ser
 210 215 220
 Ala Tyr Gly Leu Val Ala Ala
 225 230

<210> SEQ ID NO 97
 <211> LENGTH: 231
 <212> TYPE: PRT
 <213> ORGANISM: *Cyanothece sp.*

<400> SEQUENCE: 97

Met Gln Glu Leu Val Gln Arg Ser Glu Leu Asp Phe Thr Asn Pro Thr
 1 5 10 15
 Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu
 20 25 30

-continued

	165		170		175										
Val	Trp	Glu	Met	Leu	Asn	Gln	Val	Glu	Gly	Asp	Ala	Lys	Val	Leu	Gly
			180						185					190	
Met	Glu	Lys	Glu	Ala	Leu	Val	Glu	Asp	Phe	Met	Ile	Ser	Tyr	Gly	Glu
		195					200					205			
Ala	Leu	Ser	Asn	Ile	Gly	Phe	Ser	Thr	Arg	Asp	Ile	Met	Arg	Met	Ser
	210				215						220				
Ser	His	Gly	Leu	Val	Ala	Ala									
225					230										

<210> SEQ ID NO 99

<211> LENGTH: 240

<212> TYPE: PRT

<213> ORGANISM: *Synechococcus* sp.

<400> SEQUENCE: 99

Met	Asn	Val	Leu	Pro	Asn	Thr	Pro	Gln	Pro	Leu	Ala	Asp	Glu	Gly	Gly
1			5					10						15	
Thr	Thr	Leu	Asp	Tyr	Gly	Ser	Ala	Val	Tyr	Arg	Gln	Ala	Tyr	Ser	Arg
		20						25					30		
Ile	Asn	Gly	Ile	Val	Ile	Glu	Gly	Glu	Gln	Glu	Ala	His	Asp	Asn	Tyr
		35				40						45			
Leu	Lys	Leu	Ala	Glu	Met	Leu	Pro	Glu	Gly	Ala	Glu	Glu	Leu	His	Lys
	50					55					60				
Leu	Ala	Lys	Met	Glu	Leu	Arg	His	Met	Lys	Gly	Phe	Gln	Ser	Cys	Gly
65				70					75						80
Lys	Asn	Leu	Gln	Val	Glu	Pro	Asp	Arg	Glu	Phe	Ala	Arg	Thr	Phe	Phe
			85					90						95	
Ala	Pro	Leu	Arg	Asn	Asn	Phe	Gln	Lys	Ala	Ala	Ala	Ala	Gly	Asp	Leu
		100						105					110		
Val	Thr	Cys	Leu	Val	Ile	Gln	Ser	Leu	Ile	Ile	Glu	Cys	Phe	Ala	Ile
		115				120						125			
Ala	Ala	Tyr	Asn	Ile	Tyr	Ile	Pro	Val	Ala	Asp	Glu	Phe	Ala	Arg	Lys
	130				135						140				
Ile	Thr	Glu	Gly	Val	Val	Lys	Asp	Glu	Tyr	Leu	His	Leu	Asn	Phe	Gly
145				150				155						160	
Glu	Arg	Trp	Leu	Gly	Glu	His	Phe	Gly	Glu	Val	Lys	Gly	Gln	Ile	Glu
			165					170						175	
Ala	Ala	Asn	Ala	Gln	Asn	Leu	Pro	Leu	Val	Trp	Gln	Met	Leu	Gln	Gln
		180						185					190		
Val	Asp	Gln	Asp	Val	Glu	Ala	Ile	Tyr	Met	Asp	Arg	Glu	Ala	Ile	Val
		195					200				205				
Glu	Asp	Phe	Met	Ile	Ala	Tyr	Gly	Glu	Ala	Leu	Ala	Asn	Ile	Gly	Phe
	210				215						220				
Ser	Thr	Arg	Glu	Val	Met	Arg	Leu	Ser	Ala	Gln	Gly	Leu	Arg	Ala	Ala
225				230					235					240	

<210> SEQ ID NO 100

<211> LENGTH: 241

<212> TYPE: PRT

<213> ORGANISM: *Synechococcus* sp.

<400> SEQUENCE: 100

Met	Ala	Ser	Ser	Leu	Leu	Asp	Pro	Ala	Val	Asp	Gly	Thr	Pro	Val	Leu
1				5					10					15	
Asp	Val	Glu	Leu	Pro	Asp	Phe	Thr	Thr	Glu	Ala	Tyr	Lys	Ser	Ala	Tyr

-continued

20					25					30					
Ser	Arg	Ile	Asn	Ala	Ile	Val	Ile	Glu	Gly	Glu	Gln	Glu	Ala	His	Asp
	35						40					45			
Asn	Tyr	Ile	Ser	Leu	Gly	Thr	Leu	Ile	Pro	Asp	Gln	Ala	Asp	Glu	Leu
	50					55					60				
Ala	Gln	Leu	Ala	Arg	Met	Glu	Met	Lys	His	Met	Lys	Gly	Phe	Gln	Ala
	65					70					75				80
Cys	Gly	Lys	Asn	Leu	Ser	Val	Glu	Pro	Asp	Met	Val	Phe	Ala	Lys	Glu
				85					90					95	
Phe	Phe	Ser	Asp	Leu	His	Gly	Asn	Phe	Arg	Ser	Ala	Leu	Glu	Glu	Asn
			100						105					110	
Lys	Val	Val	Thr	Cys	Leu	Val	Ile	Gln	Ala	Leu	Met	Ile	Glu	Ala	Phe
			115					120					125		
Ala	Ile	Ala	Ala	Tyr	His	Ile	Tyr	Ile	Pro	Val	Ala	Asp	Pro	Phe	Ala
	130						135					140			
Arg	Lys	Ile	Thr	Glu	Gly	Val	Val	Lys	Asp	Glu	Tyr	Thr	His	Leu	Asn
	145					150					155				160
Tyr	Gly	Gln	Glu	Trp	Leu	Lys	Ala	Asn	Phe	Asp	Ser	Ser	Arg	Asp	Glu
				165					170					175	
Ile	Ile	Glu	Ala	Asn	Lys	Ala	Asn	Leu	Pro	Ile	Ile	Arg	Arg	Met	Leu
			180					185						190	
Glu	Glu	Val	Ala	Asp	Asp	Ala	Ala	Glu	Leu	Lys	Met	Glu	Lys	Glu	Ser
		195						200				205			
Leu	Ile	Glu	Asp	Phe	Leu	Ile	Ala	Tyr	Gln	Glu	Ala	Leu	Met	Asp	Ile
	210						215					220			
Gly	Phe	Asn	Ser	Arg	Asp	Leu	Ala	Arg	Met	Ser	Ala	Ala	Ala	Leu	Val
	225					230					235				240

Ala

<210> SEQ ID NO 101

<211> LENGTH: 243

<212> TYPE: PRT

<213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 101

Met	Pro	Thr	Pro	Val	Thr	Ser	Glu	Val	Ala	Val	Leu	Asp	Gly	Gln	Ala
				5					10					15	
Gly	Ser	Ala	Gln	Ala	Leu	Pro	Asp	Phe	Ser	Ser	Glu	Ala	Tyr	Lys	Asp
			20					25					30		
Ala	Tyr	Ser	Arg	Ile	Asn	Ala	Ile	Val	Ile	Glu	Gly	Glu	Gln	Glu	Ala
		35					40					45			
His	Asp	Asn	Tyr	Ile	Ser	Leu	Gly	Thr	Leu	Ile	Pro	Glu	Gln	Ala	Asp
	50					55					60				
Glu	Leu	Ala	Arg	Leu	Ala	Arg	Met	Glu	Met	Lys	His	Met	Lys	Gly	Phe
	65					70					75				80
Met	Ser	Cys	Gly	Arg	Asn	Leu	Gly	Val	Glu	Ala	Asp	Met	Pro	Phe	Ala
				85					90					95	
Lys	Glu	Phe	Phe	Gly	Pro	Leu	His	Gly	Asn	Phe	Gln	Thr	Ala	Leu	Lys
			100					105					110		
Glu	Gly	Lys	Val	Val	Thr	Cys	Leu	Leu	Ile	Gln	Ala	Leu	Leu	Ile	Glu
		115					120					125			
Ala	Phe	Ala	Ile	Ser	Ala	Tyr	His	Ile	Tyr	Ile	Pro	Val	Ala	Asp	Pro
	130						135					140			
Phe	Ala	Arg	Lys	Ile	Thr	Glu	Gly	Val	Val	Lys	Asp	Glu	Tyr	Thr	His

-continued

145	150	155	160
Leu Asn Tyr Gly Gln Glu Trp Leu Lys Ala Asn Phe Glu Ala Ser Arg	165	170	175
Glu Glu Leu Met Glu Ala Asn Lys Val Asn Leu Pro Leu Ile Arg Ser	180	185	190
Met Leu Glu Gln Val Ala Lys Asp Ala Ala Val Leu Lys Met Glu Lys	195	200	205
Glu Asp Leu Ile Glu Asp Phe Leu Ile Ala Tyr Gln Glu Ala Leu Glu	210	215	220
Glu Ile Gly Phe Thr Ser Arg Asp Ile Ala Arg Met Ala Ala Ala Ala	225	230	235
			240

Leu Ser Ile

<210> SEQ ID NO 102

<211> LENGTH: 239

<212> TYPE: PRT

<213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 102

Met Thr Thr Leu Asn Ala Pro Glu Ala Ala Val Val Glu Gly Leu Asp	1	5	10	15
Ala Leu Pro Asp Phe Thr Thr Glu Ala Tyr Lys Asp Ala Tyr Ser Arg	20	25	30	
Ile Asn Ala Ile Val Ile Glu Gly Glu Gln Glu Ala His Asp Asn Tyr	35	40	45	
Ile Ser Leu Gly Ser Leu Ile Pro Asp Gln Lys Asp Glu Leu Ala Lys	50	55	60	
Leu Ala Arg Met Glu Met Lys His Met Lys Gly Phe Thr Ser Cys Gly	65	70	75	80
Arg Asn Leu Gly Val Glu Ala Asp Met Val Phe Ala Lys Lys Phe Phe	85	90	95	
Glu Pro Leu His Gly Asn Phe Gln Ala Ala Leu Lys Glu Gly Lys Val	100	105	110	
Val Thr Cys Leu Leu Ile Gln Ala Leu Leu Ile Glu Ala Phe Ala Ile	115	120	125	
Ser Ala Tyr His Ile Tyr Ile Pro Val Ala Asp Pro Phe Ala Arg Lys	130	135	140	
Ile Thr Glu Gly Val Val Lys Asp Glu Tyr Thr His Leu Asn Tyr Gly	145	150	155	160
Gln Glu Trp Leu Lys Ala Asn Phe Glu Ala Ser Lys Asp Glu Leu Phe	165	170	175	
Glu Ala Asn Lys Ala Asn Leu Pro Leu Ile Arg Ser Met Leu Glu Glu	180	185	190	
Val Ala Ser Asp Ala Ala Val Leu His Met Glu Lys Glu Asp Leu Ile	195	200	205	
Glu Asp Phe Leu Ile Ala Tyr Gln Glu Ala Leu Gly Glu Ile Gly Phe	210	215	220	
Thr Ser Arg Asp Ile Ala Arg Met Ala Ala Ala Ala Leu Ala Val	225	230	235	

<210> SEQ ID NO 103

<211> LENGTH: 243

<212> TYPE: PRT

<213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 103

-continued

Met Pro Thr Pro Val Thr Ser Glu Val Ala Val Leu Asp Glu Gln Ala
 1 5 10 15
 Gly Ser Ala Ser Leu Leu Pro Asp Phe Ser Ser Glu Ala Tyr Lys Asp
 20 25 30
 Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu Gln Glu Ala
 35 40 45
 His Asp Asn Tyr Ile Ser Leu Gly Thr Leu Ile Pro Asp Gln Ala Asp
 50 55 60
 Glu Leu Ala Arg Leu Ala Arg Met Glu Met Lys His Met Lys Gly Phe
 65 70 75 80
 Thr Ser Cys Gly Arg Asn Leu Gly Val Asp Ala Asp Met Pro Phe Ala
 85 90 95
 Lys Thr Phe Phe Ala Pro Leu His Gly Asn Phe Gln Thr Ala Leu Lys
 100 105 110
 Asp Gly Lys Val Val Thr Cys Leu Leu Ile Gln Ala Leu Leu Ile Glu
 115 120 125
 Ala Phe Ala Ile Ser Ala Tyr His Ile Tyr Ile Pro Val Ala Asp Pro
 130 135 140
 Phe Ala Arg Lys Ile Thr Glu Gly Val Val Lys Asp Glu Tyr Thr His
 145 150 155 160
 Leu Asn Tyr Gly Gln Glu Trp Leu Lys Ala Asn Phe Asp Ala Ser Arg
 165 170 175
 Glu Glu Leu Met Glu Ala Asn Lys Val Asn Leu Pro Leu Ile Arg Ser
 180 185 190
 Met Leu Glu Gln Val Ala Glu Asp Ala Ala Val Leu Lys Met Glu Lys
 195 200 205
 Glu Asp Leu Ile Glu Asp Phe Leu Ile Ala Tyr Gln Glu Ala Leu Glu
 210 215 220
 Gln Ile Gly Phe Thr Ser Arg Asp Ile Ala Arg Met Ala Ala Ala Ala
 225 230 235 240
 Leu Ala Val

<210> SEQ ID NO 104

<211> LENGTH: 239

<212> TYPE: PRT

<213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 104

Met Thr Thr Leu Asn Ala Pro Glu Ala Ala Val Val Glu Gly Leu Asp
 1 5 10 15
 Ala Leu Pro Asp Phe Thr Thr Glu Ala Tyr Lys Asp Ala Tyr Ser Arg
 20 25 30
 Ile Asn Ala Ile Val Ile Glu Gly Glu Gln Glu Ala His Asp Asn Tyr
 35 40 45
 Ile Ser Leu Gly Thr Leu Ile Pro Asp Gln Lys Asp Glu Leu Ala Lys
 50 55 60
 Leu Ala Arg Met Glu Met Lys His Met Lys Gly Phe Thr Ser Cys Gly
 65 70 75 80
 Arg Asn Leu Gly Val Glu Ala Asp Leu Ala Phe Ala Lys Lys Phe Phe
 85 90 95
 Glu Pro Leu His Gly Asn Phe Gln Ala Ala Leu Lys Glu Gly Lys Val
 100 105 110
 Val Thr Cys Leu Leu Ile Gln Ala Leu Leu Ile Glu Ala Phe Ala Ile
 115 120 125

-continued

Ser Ala Tyr His Ile Tyr Ile Pro Val Ala Asp Pro Phe Ala Arg Lys
 130 135 140
 Ile Thr Glu Gly Val Val Lys Asp Glu Tyr Thr His Leu Asn Tyr Gly
 145 150 155 160
 Gln Glu Trp Leu Lys Ala Asn Phe Glu Ala Ser Lys Asp Glu Leu Phe
 165 170 175
 Glu Ala Asn Lys Ala Asn Leu Pro Leu Ile Arg Ser Met Leu Glu Asp
 180 185 190
 Val Ala Ser Asp Ala Ala Val Leu His Met Glu Lys Glu Asp Leu Ile
 195 200 205
 Glu Asp Phe Leu Ile Ala Tyr Gln Glu Ala Leu Gly Glu Ile Gly Phe
 210 215 220
 Thr Ser Arg Asp Ile Ala Arg Met Ala Ala Ala Leu Ala Val
 225 230 235

<210> SEQ ID NO 105

<211> LENGTH: 243

<212> TYPE: PRT

<213> ORGANISM: *Synechococcus* sp.

<400> SEQUENCE: 105

Met Ala Pro Ala Asn Val Leu Pro Asn Thr Pro Pro Ser Pro Thr Asp
 1 5 10 15
 Gly Gly Gly Thr Ala Leu Asp Tyr Ser Ser Pro Arg Tyr Arg Gln Ala
 20 25 30
 Tyr Ser Arg Ile Asn Gly Ile Val Ile Glu Gly Glu Gln Glu Ala His
 35 40 45
 Asp Asn Tyr Leu Lys Leu Ala Glu Met Leu Pro Glu Ala Ala Glu Glu
 50 55 60
 Leu Arg Lys Leu Ala Lys Met Glu Leu Arg His Met Lys Gly Phe Gln
 65 70 75 80
 Ala Cys Gly Lys Asn Leu Gln Val Glu Pro Asp Val Glu Phe Ala Arg
 85 90 95
 Ala Phe Phe Ala Pro Leu Arg Asp Asn Phe Gln Ser Ala Ala Ala Ala
 100 105 110
 Gly Asp Leu Val Ser Cys Phe Val Ile Gln Ser Leu Ile Ile Glu Cys
 115 120 125
 Phe Ala Ile Ala Ala Tyr Asn Ile Tyr Ile Pro Val Ala Asp Asp Phe
 130 135 140
 Ala Arg Lys Ile Thr Glu Gly Val Val Lys Asp Glu Tyr Leu His Leu
 145 150 155 160
 Asn Phe Gly Glu Arg Trp Leu Gly Glu His Phe Ala Glu Val Lys Ala
 165 170 175
 Gln Ile Glu Ala Ala Asn Ala Gln Asn Leu Pro Leu Val Arg Gln Met
 180 185 190
 Leu Gln Gln Val Glu Ala Asp Val Glu Ala Ile Tyr Met Asp Arg Glu
 195 200 205
 Ala Ile Val Glu Asp Phe Met Ile Ala Tyr Gly Glu Ala Leu Ala Ser
 210 215 220
 Ile Gly Phe Asn Thr Arg Glu Val Met Arg Leu Ser Ala Gln Gly Leu
 225 230 235 240
 Arg Ala Ala

<210> SEQ ID NO 106

-continued

<211> LENGTH: 239

<212> TYPE: PRT

<213> ORGANISM: *Synechococcus* sp.

<400> SEQUENCE: 106

Met Thr Thr Leu Asn Ala Pro Glu Ala Ala Val Val Glu Gly Leu Asp
 1 5 10 15
 Ala Leu Pro Asp Phe Thr Thr Glu Ala Tyr Lys Asp Ala Tyr Ser Arg
 20 25 30
 Ile Asn Ala Ile Val Ile Glu Gly Glu Gln Glu Ala His Asp Asn Tyr
 35 40 45
 Ile Ser Leu Gly Ser Leu Ile Pro Asp Gln Lys Asp Glu Leu Ala Lys
 50 55 60
 Leu Ala Arg Met Glu Met Lys His Met Lys Gly Phe Thr Ser Cys Gly
 65 70 75 80
 Arg Asn Leu Gly Val Glu Ala Asp Met Val Phe Ala Lys Thr Phe Phe
 85 90 95
 Glu Pro Leu His Gly Asn Phe Gln Ala Ala Leu Lys Glu Gly Lys Val
 100 105 110
 Val Thr Cys Leu Leu Ile Gln Ala Leu Leu Ile Glu Ala Phe Ala Ile
 115 120 125
 Ser Ala Tyr His Ile Tyr Ile Pro Val Ala Asp Pro Phe Ala Arg Lys
 130 135 140
 Ile Thr Glu Gly Val Val Lys Asp Glu Tyr Thr His Leu Asn Tyr Gly
 145 150 155 160
 Gln Glu Trp Leu Lys Ala Asn Phe Glu Ala Ser Lys Asp Glu Leu Phe
 165 170 175
 Glu Ala Asn Lys Ala Asn Leu Pro Leu Ile Arg Ser Met Leu Glu Glu
 180 185 190
 Val Ala Ser Asp Ala Ala Val Leu His Met Glu Lys Glu Asp Leu Ile
 195 200 205
 Glu Asp Phe Leu Ile Ala Tyr Gln Glu Ala Leu Gly Glu Ile Gly Phe
 210 215 220
 Thr Ser Arg Asp Ile Ala Arg Met Ala Ala Ala Ala Leu Ala Val
 225 230 235

<210> SEQ ID NO 107

<211> LENGTH: 239

<212> TYPE: PRT

<213> ORGANISM: *Synechococcus* sp.

<400> SEQUENCE: 107

Met Thr Thr Leu Asn Ala Pro Glu Ala Ser Val Met Glu Gly Gln Asp
 1 5 10 15
 Ala Leu Pro Asp Phe Thr Thr Glu Ala Tyr Lys Asp Ala Tyr Ser Arg
 20 25 30
 Ile Asn Ala Ile Val Ile Glu Gly Glu Gln Glu Ala His Asp Asn Tyr
 35 40 45
 Ile Ser Leu Gly Thr Leu Ile Pro Asp Gln Ala Glu Glu Leu Ala Arg
 50 55 60
 Leu Ala Arg Met Glu Met Lys His Met Lys Gly Phe Thr Ser Cys Gly
 65 70 75 80
 Arg Asn Leu Gly Val Gln Ala Asp Met Ala Phe Ala Arg Lys Phe Phe
 85 90 95
 Glu Pro Leu His Gly Asn Phe Gln Ser Ala Leu Lys Glu Gly Lys Val
 100 105 110

-continued

Val Thr Cys Leu Leu Ile Gln Ala Leu Leu Ile Glu Ala Phe Ala Ile
 115 120 125

Ser Ala Tyr His Ile Tyr Ile Pro Val Ala Asp Pro Phe Ala Arg Lys
 130 135 140

Ile Thr Glu Gly Val Val Lys Asp Glu Tyr Thr His Leu Asn Tyr Gly
 145 150 155 160

Gln Glu Trp Leu Lys Ala Asn Phe Glu Ala Ser Lys Glu Glu Leu Phe
 165 170 175

Glu Ala Asn Lys Ala Asn Leu Pro Leu Ile Arg Ser Met Leu Glu Asp
 180 185 190

Val Ala Ala Asp Ala Ala Val Leu His Met Glu Lys Glu Asp Leu Ile
 195 200 205

Glu Asp Phe Leu Ile Ala Tyr Asn Glu Ala Leu Ser Glu Ile Gly Phe
 210 215 220

Ser Ser Arg Asp Ile Ala Arg Met Ala Ala Ala Ala Leu Ala Leu
 225 230 235

<210> SEQ ID NO 108

<211> LENGTH: 242

<212> TYPE: PRT

<213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 108

Met Pro Thr Leu Asp Ser Thr Leu Val Ala Val Leu Asp Asp Gln Gln
 1 5 10 15

Gly Leu Ala Glu Leu Pro Asp Phe Thr Thr Asp Ala Tyr Lys Asp Ala
 20 25 30

Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu Lys Glu Ala His
 35 40 45

Asp Asn Tyr Leu Ser Leu Gly Thr Leu Ile Pro Glu Gln Ala Glu Glu
 50 55 60

Leu Ala Lys Leu Ala Lys Met Glu Met Lys His Met Lys Gly Phe Thr
 65 70 75 80

Ala Cys Ala Lys Asn Leu Asp Val Val Ala Asp Met Pro Phe Ala Gln
 85 90 95

Glu Phe Phe Ala Pro Leu His Gly Asn Phe Gln Ser Ala Leu Lys Glu
 100 105 110

Gly Lys Val Val Thr Cys Leu Leu Ile Gln Ala Leu Leu Ile Glu Ala
 115 120 125

Phe Ala Ile Ser Ala Tyr His Ile Tyr Ile Pro Val Ala Asp Pro Phe
 130 135 140

Ala Arg Lys Ile Thr Glu Gly Val Val Lys Asp Glu Tyr Thr His Leu
 145 150 155 160

Asn Tyr Gly Gln Glu Trp Leu Lys Ala Asn Phe Glu Ala Ser Arg Asp
 165 170 175

Glu Leu Met Glu Ala Asn Lys Val Asn Leu Pro Leu Ile Arg Ser Met
 180 185 190

Leu Glu Gln Val Ala Ala Asp Ala Ser Val Leu His Met Glu Lys Glu
 195 200 205

Asp Leu Ile Glu Asp Phe Leu Ile Ala Tyr Gln Glu Ala Leu Asn Glu
 210 215 220

Ile Gly Phe Ser Ser Arg Asp Ile Ala Arg Met Ala Ala Ala Ala Leu
 225 230 235 240

Ser Ile

-continued

<210> SEQ ID NO 109
 <211> LENGTH: 239
 <212> TYPE: PRT
 <213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 109

Met Thr Thr Leu Asn Ala Pro Asp Ala Ala Val Val Glu Gly Leu Asp
 1 5 10 15
 Ala Leu Pro Asp Phe Thr Thr Glu Ala Tyr Lys Asp Ala Tyr Ser Arg
 20 25 30
 Ile Asn Ala Ile Val Ile Glu Gly Glu Gln Glu Ala His Asp Asn Tyr
 35 40 45
 Ile Ala Leu Gly Thr Leu Ile Pro Asp Gln Lys Asp Glu Leu Ala Arg
 50 55 60
 Leu Ala Arg Met Glu Met Lys His Met Lys Gly Phe Thr Ser Cys Gly
 65 70 75 80
 Arg Asn Leu Gly Val Lys Ala Asp Met Val Phe Ala Lys Thr Phe Phe
 85 90 95
 Glu Pro Leu His Arg Asn Phe Gln Ser Ala Leu Gln Glu Gly Lys Val
 100 105 110
 Val Thr Cys Leu Leu Ile Gln Ala Leu Leu Ile Glu Ala Phe Ala Ile
 115 120 125
 Ser Ala Tyr His Ile Tyr Ile Pro Val Ala Asp Pro Phe Ala Arg Lys
 130 135 140
 Ile Thr Glu Gly Val Val Lys Asp Glu Tyr Thr His Leu Asn Tyr Gly
 145 150 155 160
 Gln Glu Trp Leu Lys Ala Asn Phe Glu Ala Ser Lys Asp Glu Leu Phe
 165 170 175
 Glu Ala Asn Lys Ala Asn Leu Pro Leu Ile Arg Ser Met Leu Asp Asp
 180 185 190
 Val Ala Gly Asp Ala Ala Val Leu His Met Glu Lys Glu Asp Leu Ile
 195 200 205
 Glu Asp Phe Leu Ile Ala Tyr Gln Glu Ala Leu Gly Glu Ile Gly Phe
 210 215 220
 Thr Ser Arg Asp Ile Ala Arg Met Ala Ala Ala Ala Leu Ala Val
 225 230 235

<210> SEQ ID NO 110
 <211> LENGTH: 253
 <212> TYPE: PRT
 <213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 110

Met Pro Ser Leu Glu Thr Thr Ile Ala Ala Ser Glu Thr Ala Ser Ala
 1 5 10 15
 Ser Ala Ser Met Ala Val Gly Gly Ser Val Glu Gln Asp Leu Gly Leu
 20 25 30
 Pro Asp Phe Ser Ser Ser Thr Tyr Lys Asp Ala Tyr Ser Arg Ile Asn
 35 40 45
 Ala Ile Val Ile Glu Gly Glu Gln Glu Ala His Asp Asn Tyr Ile Ser
 50 55 60
 Leu Gly His Leu Ile Pro Asp Gln Ala Glu Glu Leu Glu Arg Leu Ala
 65 70 75 80
 Arg Met Glu Leu Lys His Lys Lys Gly Phe Thr Ala Cys Ala Lys Asn
 85 90 95

-continued

Leu Ser Val Ile Ala Asp Met Asp Phe Ala Lys Glu Phe Phe Ser Pro
 100 105 110
 Leu His Gly Asn Phe Gln Ala Ala Leu Ala Glu Gly Lys Val Val Thr
 115 120 125
 Cys Leu Leu Ile Gln Ala Ile Leu Ile Glu Ala Phe Ala Ile Ser Ala
 130 135 140
 Tyr His Ile Tyr Ile Pro Val Ala Asp Pro Phe Ala Arg Lys Ile Thr
 145 150 155 160
 Glu Gly Val Val Lys Asp Glu Tyr Thr His Leu Asn Tyr Gly Gln Glu
 165 170 175
 Trp Leu Lys Ala Asn Leu Glu Ser Ser Arg Gly Glu Leu Glu Gln Ala
 180 185 190
 Asn Arg Val Asn Leu Pro Leu Val Arg Lys Met Leu Glu Gln Val Ala
 195 200 205
 Gly Asp Ala Ala Val Leu His Met Asp Gln Glu Asp Leu Met Ala Asp
 210 215 220
 Phe Met Thr Ser Tyr Gln Glu Ala Leu Thr Asp Ile Gly Phe Thr Thr
 225 230 235 240
 Arg Glu Ile Ala Lys Met Ala Thr Ala Ala Leu Leu Gly
 245 250

<210> SEQ ID NO 111

<211> LENGTH: 235

<212> TYPE: PRT

<213> ORGANISM: *Gloeobacter violaceus*

<400> SEQUENCE: 111

Met Asn Arg Thr Ala Pro Ser Ser Ala Ala Leu Asp Tyr Arg Ser Asp
 1 5 10 15
 Thr Tyr Arg Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Leu Glu Gly
 20 25 30
 Glu Arg Glu Ala His Ala Asn Tyr Leu Thr Leu Ala Glu Met Leu Pro
 35 40 45
 Asp His Ala Glu Ala Leu Lys Lys Leu Ala Ala Met Glu Asn Arg His
 50 55 60
 Phe Lys Gly Phe Gln Ser Cys Ala Arg Asn Leu Glu Val Thr Pro Asp
 65 70 75 80
 Asp Pro Phe Ala Arg Ala Tyr Phe Glu Gln Leu Asp Gly Asn Phe Gln
 85 90 95
 Gln Ala Ala Ala Glu Gly Asp Leu Thr Thr Cys Met Val Ile Gln Ala
 100 105 110
 Leu Ile Ile Glu Cys Phe Ala Ile Ala Ala Tyr Asn Val Tyr Ile Pro
 115 120 125
 Val Ala Asp Ala Phe Ala Arg Lys Val Thr Glu Gly Val Val Lys Asp
 130 135 140
 Glu Tyr Thr His Leu Asn Phe Gly Gln Gln Trp Leu Lys Glu Arg Phe
 145 150 155 160
 Val Thr Val Arg Glu Gly Ile Glu Arg Ala Asn Ala Gln Asn Leu Pro
 165 170 175
 Ile Val Trp Arg Met Leu Asn Ala Val Glu Ala Asp Thr Glu Val Leu
 180 185 190
 Gln Met Asp Lys Glu Ala Ile Val Glu Asp Phe Met Ile Ala Tyr Gly
 195 200 205
 Glu Ala Leu Gly Asp Ile Gly Phe Ser Met Arg Asp Val Met Lys Met

-continued

Glu Leu Ala Arg Leu Ala Arg Met Glu Met Lys His Met Lys Gly Phe
 65 70 75 80
 Thr Ser Cys Gly Arg Asn Leu Gly Val Glu Ala Asp Leu Pro Phe Ala
 85 90 95
 Lys Glu Phe Phe Ala Pro Leu His Gly Asn Phe Gln Ala Ala Leu Gln
 100 105 110
 Glu Gly Lys Val Val Thr Cys Leu Leu Ile Gln Ala Leu Leu Ile Glu
 115 120 125
 Ala Phe Ala Ile Ser Ala Tyr His Ile Tyr Ile Pro Val Ala Asp Pro
 130 135 140
 Phe Ala Arg Lys Ile Thr Glu Gly Val Val Lys Asp Glu Tyr Thr His
 145 150 155 160
 Leu Asn Tyr Gly Gln Glu Trp Leu Lys Ala Asn Phe Glu Ala Ser Lys
 165 170 175
 Asp Glu Leu Met Glu Ala Asn Lys Ala Asn Leu Pro Leu Ile Arg Ser
 180 185 190
 Met Leu Glu Gln Val Ala Ala Asp Ala Ala Val Leu Gln Met Glu Lys
 195 200 205
 Glu Asp Leu Ile Glu Asp Phe Leu Ile Ala Tyr Gln Glu Ala Leu Cys
 210 215 220
 Glu Ile Gly Phe Ser Ser Arg Asp Ile Ala Arg Met Ala Ala Ala Ala
 225 230 235 240
 Leu Ala Val

<210> SEQ ID NO 114

<211> LENGTH: 239

<212> TYPE: PRT

<213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 114

Met Thr Thr Leu Asn Ala Pro Glu Ala Pro Val Leu Glu Gly Gln Asp
 1 5 10 15
 Ala Leu Pro Asp Phe Thr Thr Ala Ala Tyr Lys Asp Ala Tyr Ser Arg
 20 25 30
 Ile Asn Ala Ile Val Ile Glu Gly Glu Gln Glu Ala His Asp Asn Tyr
 35 40 45
 Ile Ser Leu Gly Thr Leu Ile Pro Glu Gln Ala Glu Glu Leu Lys Arg
 50 55 60
 Leu Ala Arg Met Glu Met Lys His Met Lys Gly Phe Thr Ser Cys Gly
 65 70 75 80
 Arg Asn Leu Gly Val Glu Ala Asp Leu Pro Phe Ala Lys Lys Phe Phe
 85 90 95
 Glu Pro Leu His Gly Asn Phe Gln Ala Ala Leu Lys Glu Gly Lys Val
 100 105 110
 Val Thr Cys Leu Leu Ile Gln Ala Leu Leu Ile Glu Ala Phe Ala Ile
 115 120 125
 Ser Ala Tyr His Ile Tyr Ile Pro Val Ala Asp Pro Phe Ala Arg Lys
 130 135 140
 Ile Thr Glu Gly Val Val Lys Asp Glu Tyr Thr His Leu Asn Tyr Gly
 145 150 155 160
 Gln Glu Trp Leu Lys Ala Asn Phe Glu Ala Ser Lys Asn Glu Leu Phe
 165 170 175
 Glu Ala Asn Lys Ala Asn Leu Pro Leu Ile Arg Ser Met Leu Glu Asp
 180 185 190

-continued

Val Ala Ala Asp Ala Ala Val Leu His Met Glu Lys Glu Asp Leu Ile
 195 200 205
 Glu Asp Phe Leu Ile Ala Tyr Gln Glu Ala Leu Gly Glu Ile Gly Phe
 210 215 220
 Thr Ser Arg Asp Ile Ala Arg Met Ala Ala Ala Ala Leu Ala Val
 225 230 235

<210> SEQ ID NO 115
 <211> LENGTH: 243
 <212> TYPE: PRT
 <213> ORGANISM: Prochlorococcus marinus

<400> SEQUENCE: 115

Met Pro Thr Leu Glu Met Pro Glu Ala Ala Val Leu Asp Ser Thr Val
 1 5 10 15
 Gly Ser Ser Glu Ala Leu Pro Asp Phe Thr Ser Asp Ala Tyr Lys Asp
 20 25 30
 Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu Gln Glu Ala
 35 40 45
 His Asp Asn Tyr Ile Ala Ile Gly Thr Leu Leu Pro Asp His Val Glu
 50 55 60
 Glu Leu Lys Arg Leu Ala Lys Met Glu Met Arg His Lys Lys Gly Phe
 65 70 75 80
 Thr Ala Cys Gly Lys Asn Leu Gly Val Thr Ala Asp Met Asp Phe Ala
 85 90 95
 Arg Glu Phe Phe Ala Pro Leu Arg Asp Asn Phe Gln Thr Ala Leu Glu
 100 105 110
 Gln Gly Lys Thr Pro Thr Cys Leu Leu Ile Gln Ala Leu Leu Ile Glu
 115 120 125
 Ala Phe Ala Ile Ser Ala Tyr His Thr Tyr Ile Pro Val Ser Asp Pro
 130 135 140
 Phe Ala Arg Lys Ile Thr Glu Gly Val Val Lys Asp Glu Tyr Thr His
 145 150 155 160
 Leu Asn Tyr Gly Glu Ala Trp Leu Lys Ala Asn Leu Glu Ser Cys Arg
 165 170 175
 Glu Glu Leu Leu Glu Ala Asn Arg Glu Asn Leu Pro Leu Ile Arg Arg
 180 185 190
 Met Leu Asp Gln Val Ala Gly Asp Ala Ala Val Leu Gln Met Asp Lys
 195 200 205
 Glu Asp Leu Ile Glu Asp Phe Leu Ile Ala Tyr Gln Glu Ser Leu Thr
 210 215 220
 Glu Ile Gly Phe Asn Thr Arg Glu Ile Thr Arg Met Ala Ala Ala Ala
 225 230 235 240
 Leu Val Ser

<210> SEQ ID NO 116
 <211> LENGTH: 249
 <212> TYPE: PRT
 <213> ORGANISM: Cyanobium sp.

<400> SEQUENCE: 116

Met Ala Ser Val Ala His Pro Ala Val Ala Val Gln Pro Ala Thr Lys
 1 5 10 15
 Pro Ala Asp Thr Ala Ala Glu Arg Gly Asp Gly Leu Pro Asp Phe Ser
 20 25 30

-continued

Ser Asp Thr Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile
 35 40 45

Glu Gly Glu Gln Glu Ala His Asp Asn Tyr Ile Ala Leu Gly Thr Leu
 50 55 60

Ile Pro Asp Gln Ala Asp Glu Leu Ala Lys Leu Ala Arg Met Glu Leu
 65 70 75 80

Lys His Met Lys Gly Phe Thr Ala Cys Ala Asn Asn Leu Gly Val Thr
 85 90 95

Ala Asp Met Pro Phe Ala Lys Glu Phe Phe Ala Pro Leu His Gly Asn
 100 105 110

Phe Gln Arg Ala Leu Ala Glu Gly Lys Val Thr Thr Cys Leu Leu Ile
 115 120 125

Gln Ala Ile Leu Ile Glu Ala Phe Ala Ile Ser Ala Tyr His Ile Tyr
 130 135 140

Ile Pro Val Ala Asp Pro Phe Ala Arg Arg Ile Thr Glu Gly Val Val
 145 150 155 160

Lys Asp Glu Tyr Thr His Leu Asn Tyr Gly Gln Glu Trp Leu Lys Ala
 165 170 175

Asn Leu Ala Asp Val Arg Glu Glu Leu Glu Gln Ala Asn Arg Glu Asn
 180 185 190

Leu Pro Leu Val Arg Lys Met Leu Glu Gln Val Ala Gly Asp Ala Ala
 195 200 205

Val Leu Gln Met Asp Lys Glu Asp Leu Met Ala Asp Phe Leu Ser Ser
 210 215 220

Tyr Gln Glu Ala Leu Met Asp Ile Gly Phe Thr Gly Arg Glu Ile Ala
 225 230 235 240

Lys Leu Ala Ala Ala Ala Leu Val Gly
 245

<210> SEQ ID NO 117

<211> LENGTH: 243

<212> TYPE: PRT

<213> ORGANISM: Prochlorococcus marinus

<400> SEQUENCE: 117

Met Pro Thr Leu Glu Met Pro Val Ala Ala Val Leu Asp Ser Thr Val
 1 5 10 15

Gly Ser Ser Glu Ala Leu Pro Asp Phe Thr Ser Asp Arg Tyr Lys Asp
 20 25 30

Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu Gln Glu Ala
 35 40 45

His Asp Asn Tyr Ile Ala Ile Gly Thr Leu Leu Pro Asp His Val Glu
 50 55 60

Glu Leu Lys Arg Leu Ala Lys Met Glu Met Arg His Lys Lys Gly Phe
 65 70 75 80

Thr Ala Cys Gly Lys Asn Leu Gly Val Glu Ala Asp Met Asp Phe Ala
 85 90 95

Arg Glu Phe Phe Ala Pro Leu Arg Asp Asn Phe Gln Thr Ala Leu Gly
 100 105 110

Gln Gly Lys Thr Pro Thr Cys Leu Leu Ile Gln Ala Leu Leu Ile Glu
 115 120 125

Ala Phe Ala Ile Ser Ala Tyr His Thr Tyr Ile Pro Val Ser Asp Pro
 130 135 140

Phe Ala Arg Lys Ile Thr Glu Gly Val Val Lys Asp Glu Tyr Thr His
 145 150 155 160

-continued

Leu Asn Tyr Gly Glu Ala Trp Leu Lys Ala Asn Leu Glu Ser Cys Arg
 165 170 175
 Glu Glu Leu Leu Glu Ala Asn Arg Glu Asn Leu Pro Leu Ile Arg Arg
 180 185 190
 Met Leu Asp Gln Val Ala Gly Asp Ala Ala Val Leu Gln Met Asp Lys
 195 200 205
 Glu Asp Leu Ile Glu Asp Phe Leu Ile Ala Tyr Gln Glu Ser Leu Thr
 210 215 220
 Glu Ile Gly Phe Asn Thr Arg Glu Ile Thr Arg Met Ala Ala Ala Ala
 225 230 235 240
 Leu Val Ser

<210> SEQ ID NO 118
 <211> LENGTH: 239
 <212> TYPE: PRT
 <213> ORGANISM: *Synechococcus* sp.

<400> SEQUENCE: 118

Met Pro Thr Leu Asn Ala Pro Glu Val Ser Val Leu Glu Gly Gln Asp
 1 5 10 15
 Ala Leu Pro Asp Phe Thr Thr Ala Glu Tyr Lys Asp Ala Tyr Ser Arg
 20 25 30
 Ile Asn Ala Ile Val Ile Glu Gly Glu Gln Glu Ala His Asp Asn Tyr
 35 40 45
 Ile Ser Leu Gly Thr Leu Ile Pro Glu Gln Ala Asp Glu Leu Ser Arg
 50 55 60
 Leu Ala Arg Met Glu Met Lys His Met Lys Gly Phe Thr Ala Cys Ala
 65 70 75 80
 Arg Asn Leu Gly Val Glu Ala Asp Met Pro Phe Ala Lys Asp Phe Phe
 85 90 95
 Gly Pro Leu His Gly Asn Phe Gln Val Ala Leu Lys Glu Gly Lys Val
 100 105 110
 Val Thr Cys Leu Leu Ile Gln Ala Leu Leu Ile Glu Ala Phe Ala Ile
 115 120 125
 Ser Ala Tyr His Ile Tyr Ile Pro Val Ala Asp Pro Phe Ala Arg Lys
 130 135 140
 Ile Thr Glu Gly Val Val Lys Asp Glu Tyr Thr His Leu Asn Tyr Gly
 145 150 155 160
 Gln Glu Trp Leu Lys Ala Asn Phe Glu Ala Ser Lys Asp Glu Met Phe
 165 170 175
 Ala Ala Asn Lys Ala Asn Leu Pro Leu Ile Arg Ser Met Leu Glu Gly
 180 185 190
 Val Ala Ala Asp Ala Ala Val Leu His Met Glu Lys Glu Asp Leu Ile
 195 200 205
 Glu Asp Phe Leu Ile Ala Tyr Gln Glu Ala Leu Asn Glu Ile Gly Phe
 210 215 220
 Ser Ser Arg Asp Ile Ala Lys Met Ala Ala Ala Ala Leu Ala Ile
 225 230 235

<210> SEQ ID NO 119
 <211> LENGTH: 251
 <212> TYPE: PRT
 <213> ORGANISM: *Prochlorococcus marinus*

<400> SEQUENCE: 119

-continued

```

Met His Asn Glu Leu Lys Ile Thr Asp Met Gln Thr Leu Glu Ser Asn
1      5      10      15
Lys Lys Thr Ile Glu Glu Ser Thr Asn Ser Ile Ser Leu Asp Leu Pro
      20      25      30
Asp Phe Thr Thr Asp Ser Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala
      35      40      45
Ile Val Ile Glu Gly Glu Gln Glu Ala His Asp Asn Tyr Ile Ser Ile
      50      55      60
Ala Thr Leu Ile Pro Asn Glu Leu Glu Glu Leu Thr Lys Leu Ala Arg
      65      70      75      80
Met Glu Met Lys His Lys Lys Gly Phe Thr Ala Cys Gly Arg Asn Leu
      85      90      95
Asp Val Val Ala Asp Met Glu Phe Ala Lys Lys Phe Phe Ser Lys Leu
      100      105      110
His Gly Asn Phe Gln Val Ala Leu Lys Lys Gly Asn Val Thr Thr Cys
      115      120      125
Leu Leu Ile Gln Ala Ile Leu Ile Glu Ala Phe Ala Ile Ser Ala Tyr
      130      135      140
Asn Val Tyr Ile Arg Val Ala Asp Pro Phe Ala Lys Lys Ile Thr Glu
      145      150      155      160
Gly Val Val Lys Asp Glu Tyr Leu His Leu Asn Tyr Gly Gln Gln Trp
      165      170      175
Leu Lys Glu Asn Leu Ser Thr Cys Lys Asp Glu Leu Met Glu Ala Asn
      180      185      190
Lys Val Asn Leu Pro Leu Ile Lys Lys Met Leu Asp Glu Val Ala Asp
      195      200      205
Asp Ala Ser Val Leu Ala Met Asp Arg Glu Glu Leu Met Glu Glu Phe
      210      215      220
Met Ile Ala Tyr Gln Asp Thr Leu Met Glu Ile Gly Leu Asp Asn Arg
      225      230      235      240
Glu Ile Ala Arg Met Ala Met Ala Ala Ile Val
      245      250

```

<210> SEQ ID NO 120

<211> LENGTH: 229

<212> TYPE: PRT

<213> ORGANISM: Synechococcus sp.

<400> SEQUENCE: 120

```

Met Leu Glu Gly Gln Asp Ala Leu Pro Asp Phe Thr Thr Ala Glu Tyr
1      5      10      15
Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu Gln
      20      25      30
Glu Ala His Asp Asn Tyr Ile Ser Leu Gly Thr Leu Ile Pro Glu Gln
      35      40      45
Ala Glu Glu Leu Ser Arg Leu Ala Arg Met Glu Met Lys His Met Lys
      50      55      60
Gly Phe Thr Ala Cys Ala Arg Asn Leu Gly Val Glu Ala Asp Met Pro
      65      70      75      80
Phe Ala Lys Glu Phe Phe Gly Pro Leu His Gly Asn Phe Gln Val Ala
      85      90      95
Leu Lys Glu Gly Lys Val Val Thr Cys Leu Leu Ile Gln Ala Leu Leu
      100      105      110
Ile Glu Ala Phe Ala Ile Ser Ala Tyr His Ile Tyr Ile Pro Val Ala
      115      120      125

```

-continued

Asp Pro Phe Ala Arg Lys Ile Thr Glu Gly Val Val Lys Asp Glu Tyr
 130 135 140
 Thr His Leu Asn Tyr Gly Gln Glu Trp Leu Lys Ala Asn Phe Glu Ala
 145 150 155 160
 Ser Lys Asp Glu Met Phe Ala Ala Asn Lys Ala Asn Leu Pro Leu Ile
 165 170 175
 Arg Ser Met Leu Glu Gly Val Ala Ala Asp Ala Ala Val Leu His Met
 180 185 190
 Glu Lys Glu Asp Leu Ile Glu Asp Phe Leu Ile Ala Tyr Gln Glu Ala
 195 200 205
 Leu Asn Glu Ile Gly Phe Ser Ser Arg Asp Ile Ala Lys Met Ala Ala
 210 215 220
 Ala Ala Leu Ala Ile
 225

<210> SEQ ID NO 121

<211> LENGTH: 251

<212> TYPE: PRT

<213> ORGANISM: Prochlorococcus marinus

<400> SEQUENCE: 121

Met His Asn Glu Leu Lys Ile Thr Asp Met Gln Thr Leu Glu Ser Asn
 1 5 10 15
 Lys Lys Thr Ile Glu Glu Ser Ile Asn Pro Ile Ser Leu Asp Leu Pro
 20 25 30
 Asp Phe Thr Thr Asp Ser Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala
 35 40 45
 Ile Val Ile Glu Gly Glu Gln Glu Ala His Asp Asn Tyr Ile Ser Ile
 50 55 60
 Ala Thr Leu Ile Pro Asn Glu Val Glu Glu Leu Thr Lys Leu Ala Arg
 65 70 75 80
 Met Glu Met Lys His Lys Lys Gly Phe Thr Ala Cys Gly Arg Asn Leu
 85 90 95
 Gly Val Val Ala Asp Met Asp Phe Ala Lys Lys Phe Phe Ser Lys Leu
 100 105 110
 His Gly Asn Phe Gln Val Ala Leu Glu Lys Gly Asn Leu Thr Thr Cys
 115 120 125
 Leu Leu Ile Gln Ala Ile Leu Ile Glu Ala Phe Ala Ile Ser Ala Tyr
 130 135 140
 Asn Val Tyr Ile Arg Val Ala Asp Pro Phe Ala Lys Lys Ile Thr Glu
 145 150 155 160
 Gly Val Val Lys Asp Glu Tyr Leu His Leu Asn Tyr Gly Gln Glu Trp
 165 170 175
 Leu Lys Glu Asn Leu Ser Thr Cys Lys Glu Glu Leu Met Glu Ala Asn
 180 185 190
 Lys Val Asn Leu Pro Leu Ile Lys Lys Met Leu Asp Glu Val Ala Asp
 195 200 205
 Asp Ala Ser Val Leu Ala Met Asp Lys Glu Glu Leu Met Glu Glu Phe
 210 215 220
 Met Ile Ala Tyr Gln Asp Thr Leu Met Glu Ile Gly Leu Asp Asn Arg
 225 230 235 240
 Glu Ile Ala Arg Met Ala Met Ala Ala Ile Val
 245 250

-continued

```

<210> SEQ ID NO 122
<211> LENGTH: 238
<212> TYPE: PRT
<213> ORGANISM: Prochlorococcus marinus

<400> SEQUENCE: 122

Met  Gln  Thr  Leu  Glu  Ser  Asn  Lys  Lys  Thr  Asn  Leu  Glu  Asn  Ser  Ile
 1          5          10          15
Asp  Leu  Pro  Asp  Phe  Thr  Thr  Asp  Ser  Tyr  Lys  Asp  Ala  Tyr  Ser  Arg
          20          25          30
Ile  Asn  Ala  Ile  Val  Ile  Glu  Gly  Glu  Gln  Glu  Ala  His  Asp  Asn  Tyr
          35          40          45
Ile  Ser  Leu  Ala  Thr  Leu  Ile  Pro  Asn  Glu  Leu  Glu  Glu  Leu  Thr  Lys
          50          55          60
Leu  Ala  Lys  Met  Glu  Leu  Lys  His  Lys  Arg  Gly  Phe  Thr  Ala  Cys  Gly
 65          70          75          80
Arg  Asn  Leu  Gly  Val  Gln  Ala  Asp  Met  Ile  Phe  Ala  Lys  Glu  Phe  Phe
          85          90          95
Ser  Lys  Leu  His  Gly  Asn  Phe  Gln  Val  Ala  Leu  Ser  Asn  Gly  Lys  Thr
          100          105          110
Thr  Thr  Cys  Leu  Leu  Ile  Gln  Ala  Ile  Leu  Ile  Glu  Ala  Phe  Ala  Ile
          115          120          125
Ser  Ala  Tyr  His  Val  Tyr  Ile  Arg  Val  Ala  Asp  Pro  Phe  Ala  Lys  Lys
          130          135          140
Ile  Thr  Gln  Gly  Val  Val  Lys  Asp  Glu  Tyr  Leu  His  Leu  Asn  Tyr  Gly
 145          150          155          160
Gln  Glu  Trp  Leu  Lys  Glu  Asn  Leu  Ala  Thr  Cys  Lys  Asp  Glu  Leu  Met
          165          170          175
Glu  Ala  Asn  Lys  Val  Asn  Leu  Pro  Leu  Ile  Lys  Lys  Met  Leu  Asp  Gln
          180          185          190
Val  Ser  Glu  Asp  Ala  Ser  Val  Leu  Ala  Met  Asp  Arg  Glu  Glu  Leu  Met
          195          200          205
Glu  Glu  Phe  Met  Ile  Ala  Tyr  Gln  Asp  Thr  Leu  Leu  Glu  Ile  Gly  Leu
          210          215          220
Asp  Asn  Arg  Glu  Ile  Ala  Arg  Met  Ala  Met  Ala  Ala  Ile  Val
 225          230          235

```

```

<210> SEQ ID NO 123
<211> LENGTH: 242
<212> TYPE: PRT
<213> ORGANISM: Prochlorococcus marinus

<400> SEQUENCE: 123

Met  Pro  Thr  Leu  Glu  Ser  Ser  Glu  Val  Ala  Val  Ile  Ser  Asp  Leu  Glu
 1          5          10          15
Gly  Arg  Asp  Gly  Ser  Leu  Pro  Asp  Phe  Thr  Thr  Glu  Gln  Tyr  Lys  Asp
          20          25          30
Ala  Tyr  Ser  Arg  Ile  Asn  Ala  Ile  Val  Ile  Glu  Gly  Glu  Lys  Glu  Ala
          35          40          45
His  Asp  Asn  Tyr  Val  Ala  Ile  Gly  Thr  Val  Ile  Pro  Glu  Lys  Ala  Asp
          50          55          60
Glu  Leu  Lys  Lys  Leu  Ala  Ile  Met  Glu  Leu  Arg  His  Met  Lys  Gly  Phe
 65          70          75          80
Thr  Ala  Cys  Gly  Lys  Asn  Leu  Gly  Val  Val  Ala  Asp  Met  Glu  Phe  Ala
          85          90          95
Gln  Arg  Phe  Phe  Ala  Pro  Leu  His  Gly  Asn  Phe  Gln  Lys  Ala  Leu  Glu

```

-continued

100					105					110					
Asn	Gly	Lys	Ile	Thr	Thr	Cys	Phe	Leu	Ile	Gln	Ala	Ile	Leu	Ile	Glu
	115						120					125			
Ala	Phe	Ala	Ile	Ser	Ala	Tyr	His	Val	Tyr	Ile	Arg	Val	Ala	Asp	Pro
	130					135					140				
Phe	Ala	Lys	Lys	Ile	Thr	Glu	Gly	Val	Val	Lys	Asp	Glu	Tyr	Leu	His
	145					150					155				160
Leu	Asn	Tyr	Gly	Gln	Glu	Trp	Leu	Lys	Ala	Asn	Leu	Ala	Thr	Cys	Lys
				165					170					175	
Asp	Glu	Leu	Ile	Ala	Ala	Asn	Lys	Glu	Asn	Leu	Pro	Leu	Ile	Asn	Ser
		180						185					190		
Met	Leu	Asp	Gln	Val	Ala	Asn	Asp	Ala	Gln	Val	Leu	Tyr	Met	Glu	Lys
		195					200						205		
Glu	Glu	Leu	Met	Glu	Glu	Phe	Met	Ile	Ala	Tyr	Gln	Asp	Ser	Leu	Met
	210					215					220				
Glu	Ile	Gly	Leu	Asp	Ala	Arg	Glu	Ile	Ala	Arg	Met	Ala	Leu	Ala	Ala
	225					230					235				240

Ile Ala

<210> SEQ ID NO 124

<211> LENGTH: 241

<212> TYPE: PRT

<213> ORGANISM: Prochlorococcus marinus

<400> SEQUENCE: 124

Met	Gln	Ala	Phe	Ala	Ser	Asn	Asn	Leu	Thr	Val	Glu	Lys	Glu	Glu	Leu
1				5					10					15	
Ser	Ser	Asn	Ser	Leu	Pro	Asp	Phe	Thr	Ser	Glu	Ser	Tyr	Lys	Asp	Ala
			20					25					30		
Tyr	Ser	Arg	Ile	Asn	Ala	Val	Val	Ile	Glu	Gly	Glu	Gln	Glu	Ala	Tyr
		35					40					45			
Ser	Asn	Phe	Leu	Asp	Leu	Ala	Lys	Leu	Ile	Pro	Glu	His	Ala	Asp	Glu
	50					55					60				
Leu	Val	Arg	Leu	Gly	Lys	Met	Glu	Lys	Lys	His	Met	Asn	Gly	Phe	Cys
	65					70					75				80
Ala	Cys	Gly	Arg	Asn	Leu	Ala	Val	Lys	Pro	Asp	Met	Pro	Phe	Ala	Lys
				85					90					95	
Thr	Phe	Phe	Ser	Lys	Leu	His	Asn	Asn	Phe	Leu	Glu	Ala	Phe	Lys	Val
			100						105					110	
Gly	Asp	Thr	Thr	Thr	Cys	Leu	Leu	Ile	Gln	Cys	Ile	Leu	Ile	Glu	Ser
		115						120					125		
Phe	Ala	Ile	Ser	Ala	Tyr	His	Val	Tyr	Ile	Arg	Val	Ala	Asp	Pro	Phe
	130					135						140			
Ala	Lys	Arg	Ile	Thr	Glu	Gly	Val	Val	Gln	Asp	Glu	Tyr	Leu	His	Leu
	145					150					155				160
Asn	Tyr	Gly	Gln	Glu	Trp	Leu	Lys	Ala	Asn	Leu	Glu	Thr	Val	Lys	Lys
				165					170					175	
Asp	Leu	Met	Arg	Ala	Asn	Lys	Glu	Asn	Leu	Pro	Leu	Ile	Lys	Ser	Met
		180						185					190		
Leu	Asp	Glu	Val	Ser	Asn	Asp	Ala	Glu	Val	Leu	His	Met	Asp	Lys	Glu
		195					200					205			
Glu	Leu	Met	Glu	Glu	Phe	Met	Ile	Ala	Tyr	Gln	Asp	Ser	Leu	Leu	Glu
	210					215					220				

Ile Gly Leu Asp Asn Arg Glu Ile Ala Arg Met Ala Leu Ala Ala Val

-continued

Ala Val Met Glu Leu Lys His Met Arg Gly Phe Thr Ala Cys Gly Lys
65 70 75 80

Asn Leu Gly Val Lys Ala Asp Ile Pro Phe Ala Glu Lys Phe Phe Ser
85 90 95

Pro Leu His Gly Asn Phe Gln Lys Ala Phe Lys Glu Glu Asn Leu Thr
100 105 110

Thr Cys Phe Leu Ile Gln Ala Ile Leu Ile Glu Ala Phe Ala Ile Ser
115 120 125

Ala Tyr His Val Tyr Ile Arg Val Ala Asp Pro Phe Ala Lys Lys Ile
130 135 140

Thr Glu Asn Val Val Lys Asp Glu Tyr Leu His Leu Asn Tyr Gly Gln
145 150 155 160

Gln Trp Leu Lys Ala Asn Leu Asp Thr Cys Lys Glu Glu Leu Met Lys
165 170 175

Ala Asn Lys Glu Asn Leu Pro Leu Ile Lys Ser Met Leu Asp Gln Val
180 185 190

Ala Asp Asp Ala Cys Ser Leu Ser Met Asp Lys Glu Glu Leu Met Glu
195 200 205

Glu Phe Met Ile Ala Tyr Gln Asp Ser Leu Leu Glu Ile Gly Leu Asp
210 215 220

Ser Arg Glu Ile Ala Arg Met Ala Leu Ala Ala Leu Val
225 230 235

<210> SEQ ID NO 127

<211> LENGTH: 243

<212> TYPE: PRT

<213> ORGANISM: *Prochlorococcus marinus*

<400> SEQUENCE: 127

Met Gln Thr Leu Glu Ser Asn Lys Asn Ile Gln Ile Gly Ser Ser Pro
1 5 10 15

Glu Ser Asp Ser Ala Asn Leu Pro Asp Phe Thr Thr Asp Ala Tyr Lys
20 25 30

Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu Gln Glu
35 40 45

Ala Tyr Asp Asn Tyr Ile Ser Ile Ala Thr Leu Leu Pro Asn Asp Ser
50 55 60

Glu Glu Leu Thr Lys Leu Ala Lys Met Glu Leu Lys His Lys Arg Gly
65 70 75 80

Phe Thr Ala Cys Gly Lys Asn Leu Gly Val Glu Ala Asp Met Ser Phe
85 90 95

Ala Lys Glu Phe Phe Ser Lys Leu His Gly Asn Phe Gln Ala Ala Leu
100 105 110

Lys Asn Glu Ser Leu Thr Thr Cys Leu Leu Ile Gln Ala Ile Leu Ile
115 120 125

Glu Ala Phe Ala Ile Ser Ala Tyr His Val Tyr Ile Arg Val Ala Asp
130 135 140

Pro Phe Ala Lys Lys Ile Thr Gln Gly Val Val Asn Asp Glu Tyr Leu
145 150 155 160

His Leu Asn Tyr Gly Glu Lys Trp Leu Lys Glu Asn Leu Ser Thr Cys
165 170 175

Lys Asp Glu Leu Ile Ala Ala Asn Lys Val Asn Leu Pro Ile Ile Lys
180 185 190

Lys Met Leu Asp Gln Val Ala Asp Asp Ala Ala Thr Leu Ala Met Asp
195 200 205

-continued

Lys Glu Glu Leu Met Glu Glu Phe Met Ile Ala Tyr Gln Asp Ala Leu
 210 215 220

Leu Glu Met Gly Leu Asp Asn Arg Glu Ile Ala Arg Met Ala Met Ala
 225 230 235 240

Ala Ile Val

<210> SEQ ID NO 128
 <211> LENGTH: 250
 <212> TYPE: PRT
 <213> ORGANISM: Prochlorococcus marinus

<400> SEQUENCE: 128

Met Asn Lys Ser Leu Thr Asp Met Gln Thr Leu Glu Ser Lys Lys Asp
 1 5 10 15

Ile Gln Leu Glu Gly Ser Thr Asp Asn Asp Ser Ala Asn Leu Pro Asp
 20 25 30

Phe Thr Thr Asp Ala Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile
 35 40 45

Val Ile Glu Gly Glu Gln Glu Ala Tyr Asp Asn Tyr Ile Ser Ile Ala
 50 55 60

Thr Leu Leu Pro Asn Asp Ser Glu Glu Leu Thr Lys Leu Ala Lys Met
 65 70 75 80

Glu Leu Lys His Lys Arg Gly Phe Thr Ala Cys Gly Lys Asn Leu Gly
 85 90 95

Val Glu Ala Asp Met Pro Phe Ala Lys Glu Phe Phe Ser Lys Leu His
 100 105 110

Gly Asn Phe Gln Ile Ala Leu Lys Asp Gly Asn Leu Thr Thr Cys Leu
 115 120 125

Leu Ile Gln Ala Ile Leu Ile Glu Ala Phe Ala Ile Ser Ala Tyr His
 130 135 140

Val Tyr Ile Arg Val Ala Asp Pro Phe Ala Lys Lys Ile Thr Gln Gly
 145 150 155 160

Val Val Asn Asp Glu Tyr Leu His Leu Asn Tyr Gly Glu Lys Trp Leu
 165 170 175

Lys Glu Asn Leu His Thr Cys Lys Asp Glu Leu Ile Ala Ala Asn Lys
 180 185 190

Val Asn Leu Pro Leu Ile Lys Lys Met Leu Asp Gln Val Ala Glu Asp
 195 200 205

Ala Ala Thr Leu Ser Met Asp Lys Glu Glu Leu Met Glu Glu Phe Met
 210 215 220

Ile Ala Tyr Gln Asp Ala Leu Leu Glu Met Gly Leu Asp Asn Arg Glu
 225 230 235 240

Ile Ala Arg Met Ala Met Ala Ala Ile Val
 245 250

What is claimed is:

1. A fuel composition comprising a mixture of alkanes, wherein the mixture comprises pentadecane, hexadecane, and heptadecane, wherein heptadecane and pentadecane are predominant in the mixture, and wherein at least a portion of the carbon used as raw material of the alkanes in the mixture is inorganic carbon.

2. The fuel composition of claim 1, wherein the fuel composition further comprises nonadecane.

3. The fuel composition of claim 1, wherein the fuel composition further comprises tridecane.

4. The fuel composition of claim 1, wherein the fuel composition further comprises tetradecane.

5. The fuel composition of claim 1, wherein the fuel composition is a low-sulfur fuel composition.

6. The fuel composition of claim 1, wherein the fuel composition is a carbon-neutral fuel composition.

7. The fuel composition of claim 1, wherein the fuel composition has a higher δ_p than a comparable fuel composition made from fixed atmospheric carbon or plant-derived biomass.

281

8. The fuel composition of claim 1, wherein the fuel composition further comprises diesel.

9. The fuel composition of claim 1, wherein the inorganic carbon is carbon dioxide.

10. A fuel composition comprising a mixture of alkanes, wherein the mixture comprises pentadecane, hexadecane, and heptadecane, wherein heptadecane and pentadecane are predominant in the mixture, and wherein at least a portion of the carbon in the alkanes in the mixture is inorganic carbon.

11. The fuel composition of claim 10, wherein the fuel composition further comprises nonadecene.

12. The fuel composition of claim 10, wherein the fuel composition further comprises tridecane.

13. The fuel composition of claim 10, wherein the fuel composition further comprises tetradecane.

14. The fuel composition of claim 10, wherein the fuel composition is a low-sulfur fuel composition.

15. The fuel composition of claim 10, wherein the fuel composition is a carbon-neutral fuel composition.

282

16. The fuel composition of claim 10, wherein the fuel composition has a higher δ_p than a comparable fuel composition made from fixed atmospheric carbon or plant-derived biomass.

17. The fuel composition of claim 10, wherein the fuel composition further comprises diesel.

18. The fuel composition of claim 10, wherein the inorganic carbon is derived from carbon dioxide.

19. A fuel composition comprising a mixture of alkanes, wherein the mixture comprises pentadecane, hexadecane, and heptadecane, wherein heptadecane and pentadecane are predominant in the mixture, and wherein the fuel composition has a higher δ_p than a comparable fuel composition made from fixed atmospheric carbon or plant-derived biomass.

20. The fuel composition of claim 19, wherein at least a portion of the carbon used as raw material of the alkanes in the mixture is inorganic carbon; or wherein at least a portion of the carbon in the alkanes in the mixture is inorganic carbon.

* * * * *